

10/04/03 418

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 ; Search time 59.995 Seconds  
(without alignments)  
3772.184 Million cell updates/sec

Title: US-10-043-418-1

Perfect score: 4630

Sequence: 1 MATVGVLLCLCLCLFAPRL.....IGGLSLVGGEEFELKVMSY 877

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_23:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phase:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4275	92.3	879	10	Q91LY2 hordeum vul
2	2451	52.9	902	10	Q91YF8
3	2446	52.8	902	10	O22444
4	2193.5	47.4	910	10	Q8VWV9
5	2100	45.4	907	10	Q92P26
6	2100	45.4	915	10	Q92P26
7	2089	45.1	916	10	Q9AVC3
8	2049.5	44.3	929	10	Q91GCE6
9	2015	43.5	928	10	Q91LEC9
10	1950.5	42.1	935	10	Q92P04
11	1904	41.1	855	10	Q91ZT7
12	1529	33.0	932	13	Q73626
13	1443	31.2	992	3	Q9UV08
14	1439	31.1	953	11	Q8BG16
15	1438	31.1	953	11	Q91Z45
16	1412	30.5	937	6	Q9MYM4

17	1395.5	30.1	952	4	Q8IWE7
18	1390.5	30.0	873	13	073632
19	1378.5	29.8	955	3	Q9C1S7
20	1306.5	28.2	1734	4	Q8TE24
21	1302	28.1	995	3	Q9URX4
22	1192.5	25.8	1743	5	Q19004
23	1167.5	25.2	920	5	Q9NPF8
24	1158	25.0	955	5	Q21750
25	1002	21.6	856	5	Q20722
26	954	20.6	925	4	Q8TE14
27	950.5	20.5	914	4	Q8IZM5
28	947.5	20.5	914	4	Q8IZM4
29	929.5	20.1	769	4	Q8IWM0
30	925	20.0	921	10	Q9FN05
31	913.5	19.7	966	4	Q9P0X0
32	911.5	19.7	944	4	Q14697
33	904.5	19.5	944	11	Q8BHN3
34	903.5	19.5	966	11	Q08794
35	895	19.3	779	16	Q8YV00
36	882	19.0	653	11	Q8BVM0
37	879	19.0	653	6	Q9BE70
38	878.5	19.0	944	6	P79403
39	858.5	18.5	991	10	Q93Y12
40	858	18.5	728	2	Q9RH22
41	853	18.4	910	5	Q9U3F8
42	853	18.4	924	5	Q20239
43	846.5	18.3	751	16	Q8RDL1
44	837	18.1	746	16	Q8XIN9
45	829.5	17.9	959	10	Q9LUG2

ALIGNMENTS

RESULT 1

Q9LLY2	PRELIMINARY;	PRT;	879	AA.
ID	Q9LLY2			
AC	Q9LLY2;			
DT	01-OCT-2000 (TREMREL. 15, Created)			
DT	01-OCT-2000 (TREMREL. 15, Last sequence update)			
DT	01-MAR-2003 (TREMREL. 23, Last annotation update)			
DE	High pI alpha-glucosidase.			
GN	AGL97.			
OS	Hordeum vulgare (Barley).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;			
OC	Triticeae; Hordeum.			
OX	NCBI_TaxID=4513;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. Igr;			
RX	MEDLINE=20267959; PubMed=10806244;			
RA	Frandsen T.P., Lok F., Mirgorodskaya E., Roepstorff P., Svensson B.;			
RT	"Purification, enzymatic characterization, and nucleotide sequence of			
RL	a high-isoelectric-point alpha-glucosidase from barley malt.";			
DR	EMBL; AF118226; AAF76254.1; -			
DR	InterPro; IPR000322; Glyco_hydro_31.			
DR	InterPro; IPR002052; N6_Mtase.			
DR	Pfam; PF01055; Glyco_hydro_31; 1.			
DR	PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.			
DR	PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.			
DR	PROSITE; PS00092; N6_MTASE_1.			
SQ	SEQUENCE 879 AA; 96558 MW; AF9235CE7D15B44 CRC64;			

Query Match 92.3%; Score 4275; DB 10; Length 879;  
Best Local Similarity 93.8%; Pred. No. 2.6e-313;  
Matches 828; Conservative 10; Mismatches 35; Indels 10; Gaps 6;

QY 1 MATVGVLLCLCLCLFAPRLCSKEEGPLAA--RTVLAVVTMEG-ALRAE--AATGGRS 55

DB 1 MATRSLLLCLCLCLFAPRLCSKEEGPLAAGGVRSVAVDDGGRRRLRAEAAATGGAS 60

```

QY 56 STGDVQRLAVASLETDSRLVRITDADHPRWEPQDIIIPRAPGVLDVLDHAPPASSAPLQ 115
DB 61 STGDVQRLAVASLETDSRLVRITDADHPRWEPQDIIIPRAPGVLDVLDHAPPASSAPLQ 120
QY 116 G-RVLSAGSDLVLTVAHSPFRFTVSRSTGDTLFTAPGLVFRDKYLEVTSALPAGRAS 174
DB 121 GSRVLSAGSDLVLTVAHSPFRFTVSRSTGDTLFTAPGLVFRDKYLEVTSALPAGRAS 180
QY 175 LYLGEHTKSSFRHNDSTFLWNADIGASVVDVNLGSHFFVYDVDPAPGTAGHVLSS 234
DB 181 LYLGEHTKSSFRHNDSTFLWNADIGASVVDVNLGSHFFVYDVDPAPGTAGHVLSS 240
QY 235 NGMDVLYGGSVTVYKIVGVLDFDFFAAGNPLAVVDQVLTQIARPAHPYMSFGHCY 294
DB 241 NGMDVLYGGSVTVYKIVGVLDFDFFAAGNPLAVVDQVLTQIARPAHPYMSFGHCY 300
QY 295 GYLNVSDLERVARYAKARIIELEVNMVTDIDYMGFKOFTLDRVNFNTAAELRPFVDRLHRN 354
DB 301 GYLNVSDLERVARYAKARIIELEVNMVTDIDYMGFKOFTLDRVNFNTAAELRPFVDRLHRN 360
QY 355 AOKVVLIDPGLRVDPIDATYGTGVRGMOQDIEKNGTNFVGNVPGDVYFDPMPHAA 414
DB 361 AOKVVLIDPGLRVDPIDATYGTGVRGMOQDIEKNGTNFVGNVPGDVYFDPMPHAA 417
QY 415 AEFWAREISLFRRTIPVDGLMIDMNEISNFYNPBPMDLDDPPYRINNDGTGRPINNKTV 474
DB 418 AEFWAREISLFRRTIPVDGLMIDMNEISNFYNPBPMDLDDPPYRINNDGTGRPINNKTV 477
QY 475 RPLAVHYGGVTEYEHNNFGLLEAPATHRALLRDTGRPPFVLSRSTFVSGRYTAYWTGD 534
DB 478 PASAVHYGGVTEYEHNNFGLLEAPATHRALLRDTGRPPFVLSRSTFVSGRYTAYWTGD 537
QY 535 NAATWGDRLYSINTMISLFGFMPIGADIQFNGNTTEELCGRWIOLGAFYFPSRDHSA 594
DB 538 NAATWGDRLYSINTMISLFGFMPIGADIQFNGNTTEELCGRWIOLGAFYFPSRDHSA 597
QY 595 IFTVRRELYLWPSVAASRKALGRYQLLPYFTYLMYEAHTGAPARPLFFSYPHDVAT 654
DB 598 IFTVRRELYLWPSVAASRKALGRYQLLPYFTYLMYEAHTGAPARPLFFSYPHDVAT 657
QY 655 YGVDRQFLLGRVLSVPLEPGTIVDAFFAGRWYRLDYSLAVATRTGKHVTLAPAD 714
DB 658 YGVDRQFLLGRVLSVPLEPGTIVDAFFAGRWYRLDYSLAVATRTGKHVTLAPAD 717
QY 715 TVNVHLTGCTILPQOSALTTSRARTAFHLLVALAEDGTASGYLFLDDGDSPEYGRSD 774
DB 718 TVNVHVAAGTILPQOSALTTSRARTAFHLLVALAEDGTASGYLFLDDGDSPEYGRSD 777
QY 775 WSMVRFNFKIPNNKGAIKVKSEVHNSVAQSRITLVISKVLMGHRSPAAKPLTVHNSA 834
DB 778 WSMVRFNFKIPNNKGAIKVKSEVHNSVAQSRITLVISKVLMGHRSPAAKPLTVHNSA 836
QY 835 EVEASSAGTRYONAGGLGVVAHTGGLSLVVGEBFELKVMSY 877
DB 837 EVEASSAGTRYONAGGLGVVAHTGGLSLVVGEBFELKVMSY 879

RESULT 2
Q9LYF8
ID Q9LYF8 PRELIMINARY; PRT: 902 AA.
AC Q9LYF8;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DE Alpha-glucosidase 1 (AT5g11720/T22P22_110).
GN T22P22_110.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid 11; Brassicales; Brassicaceae; Arabidopsie.
OX NCBI_TaxId=3702;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Dueterhoeft A.,
RA Bancroft I., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Shinn P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,
RA Banh J., Bowser L., Carninci P., Dale J.M., Hayashizaki Y., Heuan V.W.,
RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,
RA Pham C.K., Quach H.L., Sakurai T., Satou M., Southwick A., Tang C.C.,
RA Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis cDNA clones";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
RA Chan M.M., Chang C.H., Dale J.M., Hayashizaki Y., Heuan V.W.,
RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Quach H.L., Sakurai T., Satou M., Southwick A., Tang C.C.,
RA Toriumi M., Wallender E.K., Wong C., Wu H.C., Yamada K., Yu G.,
RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis ORF clones";
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL163814; CAB87690.1; -
DR EMBL; AY053414; AAK96644.1; -
DR EMBL; BT002222; AAN72233.1; -
DR InterPro; IPR000322; Glyco_hydro_31.
DR Pfam; Pfold055; Glyco_hydro_31; 1.
DR PROSITE; P500129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; P500707; GLYCOSYL_HYDROL_F31_2; 1.
SQ SEQUENCE 902 AA; 101118 MW; DB56817DAA000B3B CRC64;

Query Match 52.9%; Score 2451; DB 10; Length 902;
Best Local Similarity 55.2%; Pred. No. 9e-176;
Matches 471; Conservative 136; Mismatches 191; Indels 56; Gaps 13;

QY 59 DVORLAVASLETDSRLVRITDADHPRWEPQDIIIPRAPGVLDVLDHAPPASSAPLQ-- 116
DB 69 DIKSLNLHVSLETSRLRIRITDSSQQRWEIPETVIPRAG-----NHSFRFSTEDCGN 123
QY 117 ----RVLSAGSDLVLTVAH-ASPFRFTVSRSTGDTLFTAPG-----LVFRDKYLEV 164
DB 124 SPENNFLADPSDLVFTLHTNTTPGFSVSRSSGDIPLDTSPPSDSDSNTYFIFKQDPLQL 183
QY 165 TSALPAGRASLYCLGHTKSSERLHNDSTFLWNADIGASVVDVNLGSHFFVYDVDPAPGTAGH 222
DB 184 SSALPENRSLYGIHGTKSRFLIPGTMTLWNADIGSENPVNLGSHFFVYDVDPAPGTAGH 243
QY 223 ----POTAHGVLSSNGMDVLYGGSVTVYKIVGVLDFDFFAAGNPLAVVDQVLTQIAR 278
DB 244 GNEEAGTTHGVLLNSNGMDVKYEGHRTYNYVIGVLDLYFAGSPENWKNQYTELGR 303
QY 279 PAPMPYMSFGHCYCRYGYLVNSDLERVARVAKARIIELEVNMVTDIDYMGFKOFTLDRV 338
DB 304 PAPMPYMSFGHCYCRYGYLVNSDLERVARVAKARIIELEVNMVTDIDYMGFKOFTLDRV 363
QY 339 FTAAELRPFVDRVNFNTAAELRPFVDRVNFNTAAELRPFVDRVNFNTAAELRPFVDRV 398
DB 364 FPEDKMQSEVDTLHKNGQKYVLIDPGLGIV---DSSYGTYNRGMEADVFIRKNGEPYLGE 420
QY 399 VMPGVYFDFMHPAAAEFAWEISLFRRTIPVDGLMIDMNEISNFYNPBPMDLDDPPYRINNDGTGR 455
DB 421 VMPGVYFDFMHPAAAEFAWEISLFRRTIPVDGLMIDMNEISNFYNPBPMDLDDPPYRINNDGTGR 479
QY 456 PPYRINNDGTGRPINNKTVRPLAVHYGGVTEYEHNNFGLLEAPATHRALLRDTGRPPFV 515
DB 478 PPYRINNDGTGRPINNKTVRPLAVHYGGVTEYEHNNFGLLEAPATHRALLRDTGRPPFV 534

```

```

Db 480 PPYKINNSGDKRPINNKTVPATSIHFGNISSEYDAHNLGLLEAKATHQAVVDITGKRPI 539
QY 516 LSRSTFVSGRYTAYTGDNAATWGLRLYSINTMLSGFLGPMIGADICGFGNTTEEL 575
Db 540 LSRSTFVSSGKYTAHTWGDNAAKWEDLAVSIPGILNFGLPFIWVGADICGFSHTTEEL 599
QY 576 CGRWIOLGAFYPSRDSHSAIFTVRRRELYLWPSVAASGRKALGLRYQLLPFYFTLMYEAHM 635
Db 600 CRRWIOLGAFYPPARDHSSLGTAQBELYLWDSVASSARKVLGLRMLLPHLYTLMYEAHV 659
QY 636 TGAPIARLFFSYPHOVATYGVDRFOLLGRGVLSVPLEPGPTTVDAYPPAGRWYLYDY 695
Db 660 SGNPIARLFFSPQDTKTYEIDSQFLGKSIWSPALKQGAVADAYPPAGNWFOLFNY 719
QY 696 SLAVATRTGKHVRLPAPADTVNVHLTGTTILPQQSALTTTSRARRTAFLHLLVALAEDGTA 755
Db 720 SFAVGDSGKHVRLDTPADHNVHVREGSIVAMQGEALTTRDARKTPYQLLVVASRLNI 779
QY 756 SGYFLDDGDSPEYGR---RSDMSWVRPNYKIPNNKGAIKVSEVNVNSYAQSRTLVISK 812
Db 780 SGELFLDDGELNRMGAGGNRDWTLVKFRCYVTGK--SVVLRSEVNVPEYASKMKWSIGK 837
QY 813 VVLMCHRSAPAPKLLTVHNSAE-----VEASSNAGTRYQNAGGLGVHAHIGLS 862
Db 838 VTPVGFENVNK--TYEVRTSERLSRISLIKTVDNDPRFLS-----VEVSKLS 888
QY 863 LVVGEEPELKVAMS 876
Db 889 LLVGKKFEMRLRLT 902

RESULT 3
O22444 PRELIMINARY; PRT; 902 AA.
AC O22444;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Alpha-glucosidase 1 (EC 3.2.1.20).
GN AGLU1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Landberg;
RX MEDLINE=98045620; PubMed=9411456;
RA Monroe J.D., Hall B.D., Gough C.M., Stephen A.L.;
RT "Nucleotide sequence of an alpha-glucosidase gene (Accession No.
RT AF014806) from Arabidopsis thaliana (PGR97-141).";
RL Plant Physiol 115:863-863(1997).
DR EMBL; AF014806; AAB82656.1; -.
DR InterPro; IPR000322; Glyco_hydro_31.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 902 AA; 101106 MW; 71F2A89F4D8D22EA CRC64;

Query Match 52.8%; Score 2446; DB 10; Length 902;
Best Local Similarity 55.0%; Pred. No. 2.2e-175;
Matches 470; Conservative 136; Mismatches 192; Indels 56; Gaps 13;

QY 59 DVORLAVASLETSRLVRITADHPREHVEVPDIIIPRAPGDVLHDAPPASAPIQG-- 116
Db 69 DIKSLHLVSLTSESLRIRITSSQORWEIPETVIPRAG-----NHSPPRFSTEEDGDN 123
QY 117 ----RVLSPPAGSDLVLTWH-ASPFRTVSRSGTDLFTDPAG-----LVFRDKYLEV 164
Db 124 SPENFLADPSSDLVTLTNTTTPFGVSRRSSGDIILFDTSPSSDSNTYFIKQDFLOL 183

```

```

QY 165 TSALPAGRASLYGLCEHTKSGFRLRHNDSTFLMNADIGASYVDVNLVYSHGFYMDVRA-- 222
Db 184 SSALPENRNSLYGIEHTKSRFRLLIPGRTWLWADTSGENPDVNLVYSHGFYMDVRGSK 243
QY 223 ----PGTAHGVLLSSNGMDVLYGSGSYVYKIVGGVLDFFYFAGNPLAVVDQYTQLIAR 278
Db 244 GNEZAGTTHGVLLNSNGMDVKYEGHRTYVNVIGVLDLYVFAFGSPBMVWNOYTTELIGR 303
QY 279 PAPMEYWGFGHQCRCYGLNVSDILERVVARAKARIPLVEMWTDIDYMDGKDFDPLDRVN 338
Db 304 PAPMEYWGFGHQCRCYGLNVSDILEYVYDYGAKAGIPLVEMWTDIDYMDGKDFDPLDRVN 363
QY 339 FTAABLRFEVDELHRNAQKYVLIILDPGIRVDPIDATYCTFVRGMOODIFLKRNGTNFVGN 398
Db 364 FPEQKQSFVDTLHKNGQKYVLIILDPGLGV--DSSYGTNRGMEADVFIKXGEPYIGE 420
QY 399 VMPGVYPPDFMHPAAAEFWARBEISLFRRTIPVGLWIDMNEISNFYNPRPM---NALDD 455
Db 421 VMPGVYPPDFLNPAAATFWSNEIKMFOEILPLDGLWIDMNEISNFI--SPLSSGSSLDD 479
QY 456 PPYRINNDGTGRPINNKTVRPLAVHYGCVTEYBEHNLFGLEEARATGRGLRDTGRPRFV 515
Db 480 PPYKINNSGDKRPINNKTVPATSIHFGNISSEYDAHNLGLLEAKATHQAVVDITGKRPI 539
QY 516 LSRSTFVSGRYTAYTGDNAATWGLRLYSINTMLSGFLGPMIGADICGFGNTTEEL 575
Db 540 LSRSTFVSSGKYTAHTWGDNAAKWEDLAVSIPGILNFGLPFIWVGADICGFSHTTEEL 599
QY 576 CGRWIOLGAFYPSRDSHSAIFTVRRRELYLWPSVAASGRKALGLRYQLLPFYFTLMYEAHM 635
Db 600 CRRWIOLGAFYPPARDHSSLGTAQBELYLWDSVASSARKVLGLRMLLPHLYTLMYEAHV 659
QY 636 TGAPIARLFFSYPHOVATYGVDRFOLLGRGVLSVPLEPGPTTVDAYPPAGRWYLYDY 695
Db 660 SGNPIARLFFSPQDTKTYEIDSQFLGKSIWSPALKQGAVADAYPPAGNWFOLFNY 719
QY 696 SLAVATRTGKHVRLPAPADTVNVHLTGTTILPQQSALTTTSRARRTAFLHLLVALAEDGTA 755
Db 720 SFAVGDSGKHVRLDTPADHNVHVREGSIVAMQGEALTTRDARKTPYQLLVVASRLNI 779
QY 756 SGYFLDDGDSPEYGR---RSDMSWVRPNYKIPNNKGAIKVSEVNVNSYAQSRTLVISK 812
Db 780 SGELFLDDGELNRMGAGGNRDWTLVKFRCYVTGK--SVVLRSEVNVPEYASKMKWSIGK 837
QY 813 VVLMCHRSAPAPKLLTVHNSAE-----VEASSNAGTRYQNAGGLGVHAHIGLS 862
Db 838 VTPVGFENVNK--TYEVRTSERLSRISLIKTVDNDPRFLS-----VEVSKLS 888
QY 863 LVVGEEPELKVAMS 876
Db 889 LLVGKKFEMRLRLT 902

RESULT 4
Q8VMV9 PRELIMINARY; PRT; 910 AA.
AC Q8VMV9;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Putative alpha-xylosidase.
OS Pinus pinaster (Maritime pine).
ON NCBI_TaxID=71647;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Etolated hypocotyl;
RA Sanchez M., Gianzo C., Sampedro J., Revilla G., Zarra I.;
RT "Changes in alpha-xylosidase gene expression during intact and auxin-
RT induced growth of pine hypocotyls.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.

```





```

QY 504 GVLRTDGRPFVLSRSTFVSGRYTAYWTGDNATWGLDRYSINTMLSFGLFGMPMGAD 563
Db 525 GLNVQGRPFILSRSTFVSGQYAAHWTDGQNTQSLQVSLTMLNFOIFGVPWUSD 584
QY 564 ICGPNGMTTELGRWTLQCAFPPSDHSAIFTVRRELYLWPSVAASGRKALGLRYOLL 623
Db 585 ICGFYPOPTBELCNWRTEVGAFFPSPDRHANYSPROELYQMDTVADARSNALGMYKIL 644
QY 624 PYFYLMEAHMTGAPRIAPLFPSPYHDVATYGVDRFLGRGLVSPVLEPGPTTVDAY 663
Db 645 PFLYTLNIEAHMTGAPRIAPLFPSPYTCYGNRSROLGLGSMFSPVLEQKTEVEAL 704
QY 684 FPAGRWRLYDYSIAVARTGKHVRLPAPADTVNVHLTGTITLPLQOSALTTSRRARTAF 743
Db 705 FPPGSMYHMFDMTQAVVSKNGKRVTLPLAPLNFVNVHLYQNTILPTQCGGLISKDARTTPF 764
QY 744 HLLVAL---REDGTASGYLFDGDSPEYGRSDWS-MVRFNKIPNNKCAIKVKSEVHH 799
Db 765 SLVIAFPAGASEGYATGKLYLDEDEPEMKLGNQSQSTYVDYFASVGN--GTMKMSQVKE 822
QY 800 NSYAQSRRTLVSIKVLMGHRSPAPKKLTVHVN--SAEVEASSSAGT-----RYCNAG 850
Db 823 GKALSAGKWIEKVSLGLRAGAGVSIQINGSPMTKKIEVSSKEHTYVIGLEDEENKS 882
QY 851 GLGGVAHIGLSLVVGGEEFELKVAM 875
Db 883 VM---VEVRGLEMLVGKDFNMSWKM 904

RESULT 6
Q9SY7Y PRELIMINARY; PRT; 915 AA.
AC Q9SY7Y;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Alpha-xylosidase precursor (Arl68560/F24J5_10).
GN XYL1 OR F24J5.20.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids 11; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP STRAIN=cv, Columbia;
RC Sampedro J., Steiro C., Villa T.G., Revilla G., Zarra I.;
RA "Cloning and expression pattern of an alpha-xylosidase gene from
RT Arabidopsis thaliana."
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP STRAIN=cv, Columbia;
RC Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
RA Lee J.M., Li J., Gonzalez A., Liu A., Liu K., Vaysberg M., Sakano H.,
RA Chin C., Choi E., Chou J., Altati H., Araujo R., Brooks S.,
RA Buehler E., Chao Q., Conn L., Conway A.B., Dunn P., Hansen N.,
RA Howng B., Huizar L., Khan S., Kim C., Palm C., Rowley D., Shinn P.,
RA Walker M., Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
RT "The sequence of BAC F24J5 from Arabidopsis thaliana chromosome 1."
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP STRAIN=cv, Columbia;
RC Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.

```

```

RN [4]
RP SEQUENCE FROM N.A.
RA Kim C.J., Chen H., Cheuk R., Shinn P., Bowser L., Carninci P.,
RA Chan M.M., Chang C.H., Dale J.M., Hayashizaki Y., Hsuan V.W.,
RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA Tang C.C., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G., Yuan S.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis ORF clones."
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF144078; RAD37363.1; --
DR EMBL; AC008075; RAD49987.1; --
DR EMBL; AY057482; RAD09716.1; --
DR EMBL; BT002675; AAL011591.1; --
DR InterPro; IPR000322; Glyco_hydro_31.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
KW Signal.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 124 915 ALPHA-XYLOSIDASE.
SQ SEQUENCE 915 AA; 102398 MW; 28F9610DBD7EA657 CRC64;

Query Match 45.4%; Score 2100; DB 10; Length 915;
Best Local Similarity 46.6%; Pred. No. 2.7e-149;
Matches 431; Conservative 145; Mismatches 267; Indels 82; Gaps 18;

QY 7 LLLCLCLCLFAP-----RLCSSKEEGPLAARTVLAIVATMEGAL--RAEAATGG 53
Db 14 LLLALILC-FSPSTQSYKTIGKGYRLVSI-EESP-----DGFGLGYLVQKN 58
QY 54 RSTGQVORLAVYASLETDSRLRVITDADPRWEVPODIIPRAPGVDLHDAPASSAP 113
Db 59 KIYGSIDITLFLVKHETDSRLRVHITDAKQORWEVPTNLLPREQPPQGVKIGSRSP 118
QY 114 LQGRVLSAGSDLVLTVAHSPRTVSRSTGDTLFDAPGLVFRDKYLEVTSALPAGRA 173
Db 119 ITVQGIS--GSELIFSYYTTPPTFAKVRSNHETLFTNTSSLVFKDQYLEISTSLPK-EA 175
QY 174 SLYGGEHTKSS-FRLRHNDSTLWADIGASYVDVNLVYGHSPFYMVDRAPG---TAGV 229
Db 176 SLYGGENSGANGIKVLPNEPYTLTYEDVSAINLTLDYGHSPMTMDLRNVGKAYAHAV 235
QY 230 LLLSSNGMDLVYGGSYVTVYKVGVLDFVFAGPNPLAVVDQYTLIARPAMPYMSFGF 289
Db 236 LLLNSNGMDVFRGDSLYTKVIGVDFDFITAGSPFLNVVDQYTLIGRPAMPYMSLGF 295
QY 290 HQCRYGLNVSDLERVARYAKARIPLEVMVMTDIDYMDGFKDFTLDRVNTFAAEALRPFVD 349
Db 296 HQCRWGYHNLVVEDVDVNYKKAKIPLDVIMNDDDHMDGHDKDFTLNPVAYPRAKLAFLD 355
QY 350 RLHRAQKVLILDRGIRVDPIDATYGTFTVRGMOODIFLKRNGTNFVGNWPGDYVDFDF 409
Db 356 KHKIKGMKYIINDPGIGV---NASYGTQFQAMAADVFIKYGKPFLLAQWMPGPFYVDF 412
QY 410 MHPAAAEFWAREISLFRITPVDGLWIDMNEISNFYN-----PE----- 448
Db 413 LNPKTVSWHGEIEKRFHDLVIDGLWIDMNEVSNFCSGLCTIPBGKQCPSEGGPWVCL 472
QY 449 -----PMNALDDPPYRINNDGTRPINKTKVRLAVHYGGVTEYEEHNLFLGLEARATGR 503
Db 473 DCKNITTKRWDDPPYKINATGVWAPVGFKTIATSNATHYGVREYDAHSIYGFSEIATHK 532
QY 504 GVLRTDGRPFVLSRSTFVSGRYTAYWTGDNATWGLDRYSINTMLSFGLFGMPMGAD 563
Db 533 GLNVQGRPFILSRSTFVSGQYAAHWTDGQNTQSLQVSLTMLNFOIFGVPWUSD 592
QY 564 ICGPNGMTTELGRWTLQCAFPPSDHSAIFTVRRELYLWPSVAASGRKALGLRYOLL 623
Db 593 ICGFTPQPTBELCNWRTEVGAFFPSPDRHANYSPROELYQMDTVADARSNALGMYKIL 652
QY 624 PYFYLMEAHMTGAPRIAPLFPSPYHDVATYGVDRFLGRGLVSPVLEPGPTTVDAY 683

```

```

Db      653 PFLYTLNVEAHMTGAPIARLPFFPEYTECYGNSRQFLLLGSSFMISPVLEQKTEVEAL 712
QY      684 PPAGRWRLYDYSIATVTRTKHVRLLPAPADTVNHLTGTTILPQOSALTTSRARTAF 743
Db      713 FPPGSGWYHMFMTQAVVSKNGKRVTLPLAPLNFVNVHLYQNTILPTQOGLISKDARTTP 772
QY      744 HLLVAL---AEDGTASGYLFLDDGDSPEYGRSDMS--MVRFNKYIPNNKGAIKVKSEVVH 799
Db      773 SLVIAFPAGASEGATGKLYLDELEPMKLGNCQSYTVDFYASGVN--GTMKQMSQVKE 830
QY      800 NSYAQSRLTVISKVLMGHRSPAPAKKLTVHVN--SAEVEASSAGT-----RYQNAG 850
Db      831 GKFAKSGMWIEKVSVLGLRCAGGVSETIQINGSPTMKKIEVSSKEHTYVIGLEDERENKS 890
QY      851 GLGGVAHIGGLSLVVGEFEFLKVM 875
Db      891 VM---VEVRGLEMLVGKDFNMSWK 912

RESULT 7
Q9AVC3 ID Q9AVC3 PRELIMINARY; PRT; 916 AA.
AC Q9AVC3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative alpha-glucosidase.
GN PPGLU.
OS Physcomitrella patens subsp. patens.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
OC Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
OX NCBI_TaxID=145481;
RN [1]
RP SEQUENCE FROM N.A.
RA Hiwataashi Y., Nishiyama T., Hasebe M.;
RT "Establishment of gene- and enhancer-trap systems of the moss,
RT Physcomitrella patens.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB057452; BAB39467.1; -
DR InterPro; IPR001064; Crystallin.
DR InterPro; IPR000322; Glyco_hydro_31.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
SQ SEQUENCE 916 AA; 102282 MW; AACEZAC6E440D883 CRC64;

Query Match 45.1%; Score 2089; DB 10; Length 916;
Best Local Similarity 48.1%; Pred. No. 1.8e-148;
Matches 422; Conservative 140; Mismatches 245; Indels 70; Gaps 17;

QY      48 EAATGSRSTGQVQLAVYASLETDRLRLVITDADHPWEVPQDIIIPRAGDVLHDAP 107
Db      53 ELITGTEYIGDIPSLRMIAVSDRLRHVITDSIHARMEVPQDIIIPRDPSSSLVTHVK 112
QY      108 PASSAPLQGRVLSAGSD--LVLTVHASFRTVSRSTGTDLPTAP----- 153
Db      113 ERDMEHSEG--LDPARNDRLQLSYTFVEFFGFAITRTSTGCLNSTPPIRQDSGEPAFN 170
QY      154 GLVPRDKYLEVTSALPAGRASLYLGHEHTK--SSFLRHNDSTFTLNADIGASYVDVNLG 212
Db      171 SMVFQDQYLEISTQLPRNN--SLFGIGESTRDPGLRLTRGLTYLTWATDIAAYKVDVLYG 229
QY      213 SHPFYMDVRAPCTAGHLLLSNGMDVLYGSGYVYTKVIGVLDFFYPFAGNPVLAVDQY 272
Db      230 AYPFMDITREGATHGVLMNSGMDIWWGEDMLTYHVIGVLDFFYPFAGPAPLAVIDQY 289
QY      273 TQLIARPAMPYMSFGHQRGYGLVNSDLRVRVARYAKARIPLEVMTDIDYMDGPKDF 332
Db      290 TNLICRTPMPYMSFGHQRGWYETIDEIKDVVANYKKANIPLDTIWINDIDYMDAYKDF 349
QY      333 TLDVNFVTAELRPVDFLRHNAQKYLILDPGIRVDPIDATYGTFFVGRMQODIFLKN- 391

```

```

Db      350 TFDPRYDENTVREPKELHANGQQYIIVLDPGISVG--YKNYSTLBRGLKDDIFLKNF 407
QY      392 GNFVGNVMPGDVYFPDFPMHAAEFNARRISLSLFRRTIPVDGLMIDMNEISNF----- 444
Db      408 GNNYLAQWMPGVPGVFPDFPLHPKASSWMTQBIADFPDKVFPDGLMIDMNEASNFCTGSACS 467
QY      445 -----YNPEPMNALDDPPPYRINNDGTGRPINNKTVRPLAVHYGCV 484
Db      468 FDTLILGKMGKNSDNDRCLLHCNVTSRFDPPYKINHVGTVDNLGVKTIAMTVKHNGV 527
QY      485 TEYEHNIFGLLEARATGRGVLRD--TGRPPFVLSRSTFVSGRYTAYWTGDNAAWTGDLR 543
Db      528 LEYDAHNLGLCESIATQK--TLRDVTKRPPILSRSTFVSGGAHTAHTGDKATMEDLK 586
QY      544 YSINTWLSFGLFGPMIGADICGFGNGNTTEELCGRWIOLGAFYPPSRDHSALFTVRRRLY 603
Db      587 YSIVSVINSGMFVPMVGADICGFAGNTTEELCRMMQLGAFYPSRNHAALGNTNSHEPY 646
QY      604 LWPVSAASGRKALGLRYOLLPVYTLMEAHMTGAPIARLPFFSYPHDVAITYGVDRQELL 663
Db      647 IWESVAESRKAALGLRYLLPHLYTLMEFNTKSGAPIARALFFFPKDLNTLAINDOELL 706
QY      664 GRGLVSPVLEPGTTVDAYPPAGRWRLYDYSIATVTRTKHVRLLPAPADTVNHLTGG 723
Db      707 GRSVLISPIVAEGLTSVNAYPKGTWYNLFDFSKIVS--TGERRMLPAPADSINVHVEG 764
QY      724 TILPQOSNLTTSRARTAFHLLVALAED--GTASGYLFLDDGDSPEYGRSDMS--MYRF 780
Db      765 QILPMQEARLTSAEVYKTPFTLVVVFVSADASASGKLVDSGVDIEMGIQGSSTFVQF 824
QY      781 NYKIPNNKGAIKVKSEVHNSYAQSRLTVISKVLMGHRSPAPAKKLTVHNSAEVEASS 840
Db      825 PAERSLHSGSL--VSRVIAGNVALEQGLVLSIRFLGVSGPVS-----DIVNGERIVSAE 878
QY      841 --SAGTRYONAGGLGVGAHIGGLSLVGEFEFLKVM 875
Db      879 QLSDYARLES-----LQVSGLSLLGLGRDFELRWAM 908

RESULT 8
Q9LGC6 ID Q9LGC6 PRELIMINARY; PRT; 929 AA.
AC Q9LGC6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE EST AU032739 (C53221) corresponds to a region of the predicted
DE gene.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
RT clone: P0504H10.-";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF002526; BAA99366.1; -.
DR Gramene; O9LGC6; -.
DR InterPro; IPR000322; Glyco_hydro_31.
DR Pfam; PF01055; Glyco_hydro_31; 1-.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
SQ SEQUENCE 929 AA; 102550 MW; 536E1D089D7F97BF CRC64;

Query Match 44.3%; Score 2049.5; DB 10; Length 929;
Best Local Similarity 45.9%; Pred. No. 1.8e-145;
Matches 424; Conservative 144; Mismatches 277; Indels 79; Gaps 19;

QY      10 CLCLCLFAPRLCSSKEEGLAA-----RTVLAVAVTMEGALRABAAATGGRSST--G 58

```

Db 15 CIPCLLFLTLASS--NCVFAAAPPKKSGCYKLVSLSVEHPEGGALGYLQVKQRTSTYGP 72  
Qy 59 DVQRLAVYASLEDSRLRVEITADHPRWEPVQDIIPRPAPGDLVLDHAPPASSAPLQGRV 118  
Db 73 DIPLLRLVYKHETKDRIRVOITDADKRWEPVNNLQJ-----EPAPPVGTGRITGVP 125  
Qy 119 LSPA--GSDVLVTHASPRFVTSRSTGDTLFTAPG-LVPRDKYLEVTSALPAGRAS 174  
Db 126 FAAGEYGEELVFTYGRDPFWFAVHRKSSREALFNTSCGALVFKDQYIEASLSLPRD-AA 184  
Qy 175 LYGLGEHTK--SSFRLRNDSTLWADIGASYDVNLYGSHHPYMDVRA---PGTAHGYL 230  
Db 185 LYGLGENTQPGIRLRNDPTTYITVTDISALNLTOLYGSHHPYVVDLRSRGGHGVAAVL 244  
Qy 231 LLSNGMDVLYGGSYVTKYKVGVLDPFFAGNPLAVVDQYTQLIARPAHPMPYMSFGH 290  
Db 245 LLSNGMDVYRGSYVTKYKVGGLDLYLFGSGTPLAVVDQYTSMLGRAPMPYMAFGH 304  
Qy 291 QCRYGYNVSDLRVRYAKARIPLEVMWTDIDYMDGFKDFTLDRVNFATAELRPFVDR 350  
Db 305 QCRWGYKNLSWVGVEGYRNAQIPLDVIWNDDHMDAAKDFLDPVNYPRPKLLEFLDK 364  
Qy 351 LHRNAQYVILDPGRVDPIDATYGTFFVGMQODIFLKGNTFNFGNVWPGVDVYFPDPM 410  
Db 365 IHAQGMKYIYVLDIPGIAV---NNTGYVQRMGQGVFIKLDGKPYLAQVWPGVYFPDFL 421  
Qy 411 HPAAEFPWAEISLFRTPIDVGLWIDMNEISNFYN-----PEPMNAL----- 453  
Db 422 NPGVSWIIDEVRPHDLVPVGLWIDMNEASNPCTGKCEIPTHLCPLNTTTPWVCL 481  
Qy 454 -----DPPYRINNDGTGRPINNKTRPLAVHYGVTYEYEHNLFGLEEARATOR 503  
Db 482 DCKNLWTRWDEPPYKINAGSQTAGLGFNTIATSYNGILEYNAHSLVGFESQAIATHQ 541  
Qy 504 GVLRTDGRFVLRSRFFVSGSYTAWTGDNAAWCDLRYSTINTMLSGCLFQMPMGAD 563  
Db 542 ALQGLQKRFPILTRSTFVSGSAYAAHMTGDNKGWENLRYSTINTMLNFGIFQMPMGAD 601  
Qy 564 ICGFNGNTEELCGRWITQLGAFYFSDHSAITFVRRELYLWPSVAASGRKALGLRYQLL 623  
Db 602 ICGFYPOTEELCNRLWELGAFYFSDHANFASPROELVWESVAKSARNALCMRYRL 661  
Qy 624 PYFTLMEAHMTGAPARLPFLFSYPHDVATYVDQRQLLGRGLVSPVLEPGFTTVDAY 683  
Db 662 PYLYTLNQAHLTGAAPVARPVFESFPDFTPCYGLSTQYLLGASVWVSPVLEQATSVSAM 721  
Qy 684 FBAGRWLYDYSLAVATRGKVRPAPADTVNVHLTGCTILPLOSALTTSERRTAF 743  
Db 722 FPGSWNLFDYTKVVSREGAVLDAPLNEINHVFTQILUPMORGGTISKEARATPF 781  
Qy 744 HLLVAL---AEDGTASGYLFLDDGDSPEYG--RRSDMSWVRFNKYIPNNKGAIKVSEVH 799  
Db 782 TLVAVFPFGATEABEAGVYVDDDERDEMYLAEGQATYVAF-YATVRGK-AVTVRSEVEL 839  
Qy 800 NSVAQSTLIVSKVLMCHRSAPAKKLTVHNSAEVSESSAGTRYQNA--GGLGG----- 854  
Db 840 GSYSLQGLLIEKLSVLGLEGTG--RDLAVHVDGANATATSRPYFAGAEALHGRDV 897  
Qy 855 -----VAHIGLGLSVVGEFEL 871  
Db 898 EGHKXSVWVEGGLALPLGKSTFW 921  
RESULT 9  
Q9LEC9  
ID Q9LEC9 PRELIMINARY; PRT; 928 AA.  
AC Q9LEC9  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Alpha-glucosidase (EC 3.2.1.20).  
GN MAL2.  
OS Solanum tuberosum subsp. tuberosum.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; lamiales; Solanales; Solanaceae; Solanum.  
OX NCBI\_TaxID=90692;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Record; TISSUE=tuberising stolon tip;  
RX MEDLINE=21362238; PubMed=11469591;  
RA Taylor M.A., Ross H.A., McRae D., Wright F., Viola R., Davies H.V.;  
RT "Copy-DNA cloning and characterisation of a potato alpha-glucosidase:  
RT expression in Escherichia coli and effects of down-regulation in  
RT transgenic potato.";  
RL Planta 213:258-264 (2001).  
DR EMBL: AJ277244; CAB96077.1; -;  
DR InterPro: IPR00322; Glyco\_hydro\_31.  
DR Pfam: PF01055; Glyco\_hydro\_31; 1.  
DR PROSITE: PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
KW Glycosidase; Hydrolase.  
SQ SEQUENCE 928 AA; 104687 MW; CEAC8634F6ED1820 CRC64;  
Query Match 43.5%; Score 2015; DB 10; Length 928;  
Best Local Similarity 46.3%; Pred. No. 7.1e-143;  
Matches 405; Conservative 136; Mismatches 265; Indels 68; Gaps 16;  
Qy 59 DVQRLAVYASLETSRLRVRITDADHPRWEPVQDIIPRPAPGDLVLDHAPPASSAPLQGRV 118  
Db 64 DIPNLQLVYKHETDNRILRHITDADKRWEPVNNLPRESPPSLKQTI--CKSRKGQFSL 121  
Qy 119 LSP---AGSDVLVTHASPRFVTSRSTGDTLF-----DTAPGLVFRDKYLEVTSAL 168  
Db 122 LSNQSYGNELMFSTSDPFSYKRSKNGQTLFNSSSESDSPYSNLVFKDQYLEISTKL 181  
Qy 169 PAGRASLYGLGEHTK--SSFRLRNDSTLWADIGASYDVNLYGSHHPYMDVDR---APG 224  
Db 182 PKD-ASLYGLGENTQPGIKIYPNDDPYLTITDSSINLNMMDLYGSHPMYMDLRNVGEA 240  
Qy 225 TAHGVLSSNGMDVLYGGSYVTKYKVGVLDPFFAGNPLAVVDQYTQLIARPAHPMPY 284  
Db 241 YAHAVLLMNSGMDVYRGSYVTKYKVGVLDPFFAGNPLAVVDQYTQLIARPAHPMPY 300  
Qy 285 WSFGHQCRGYNLVSDLRVRYAKARIPLEVMWTDIDYMDGFKDFTLDRVNFATAEL 344  
Db 301 WSFGHQCRWGYHNSLVIEDVIANYYKAKIPLDVINWDDHMDGKDFTLHPINYPGPKL 360  
Qy 345 RPFVRLHNAQYVILDPGRVDPIDATYGTFFVGMQODIFLKGNTFNFGNVWPGVDV 404  
Db 361 RAFLEKIIHAEGMHYIVINDPGLGV---NKSXGTQVORGLANDVFIKYGKPLAQVWFGAV 417  
Qy 405 YPDPMPHAAAEFWAREISLFRTPIDVGLWIDMNEISNFYN-----PE----- 448  
Db 418 HFPDPLNPKTVEWMDGDIRRFEHLAPIDGLWIDMNEVSNFCNGLCTIPGRICPNGTGPG 477  
Qy 449 -----PMNALDDPPYRINNDGTGRPINNKTRPLAVHYGVTYEYEHNLFGLEA 498  
Db 478 WICCLDCQVTKTKWDDPPYKINASGQAIPGYKTIATSYNGVREYDAHSYLSSET 537  
Qy 499 RATERGVRLDTRGPFFVLRSSTFVSGSRYTAYWTGDNAAWCDLRYSTINTMLSGCLGMP 558  
Db 538 IATHKGLQAIESGRPFIILTRATFVSGSRYAAHMTGDNKGWEDLYSLSTVNLNFGIGV 597  
Qy 559 MIGADICGF--NGMTTEELCGRWITQLGAFYFSDHSAITFVRRELYLWPSVAASGRKAL 616  
Db 598 MVGSDICGFYPAAPLEBELCNRWIQVGAFYFSDHSAITFVRRELYLWPSVAASGRKAL 657  
Qy 617 GLRYQLLPYFTLMEAHMTGAPARLPFLFSYPHDVATYVDQRQLLGRGLVSPVLEPG 676  
Db 658 GMRVYKLLPYLTLSYEAHKTGAPIVRPLFTFPNIPELYELSTQFLVGSNVWVSPVLEKA 717  
Qy 677 PTTVDAYFPAGRWLYDYSLAVATRGKVRPAPADTVNVHLTGCTILPLOSALTTS 736  
Db 718 KTKVSALFPFGTWSLFPMTQVITVKEPHYRSLADAPLHVNVHLVYQNTILPMQRCGLTK 777  
Qy 737 RARRTAFHLLVAL---AEDGTASGYLFLDDGDSPEYGRRSWMSVRFNXYKIPNNKGAIKV 793

```

Db 778 EARTPTFIIVAFPLGASEGVAKGNLFDDDELPEM-KLNGKSTYMDFHATTSTNGTVKI 836
Qy 794 KSEVHNSYAQSRTLVIVSKVLMGHRSPAPKULTVHNSAEVASSAGTRYQNAAGL- 852
Db 837 WSEVQESKALDKGWIEKVTVGLNGTGG--AFDILVDGSKVEDTSKLEFETEHEKFI 894
Qy 853 ---GG-----VAHIGGLSLVVGEEFELKVAMSY 877
Db 895 KLEGGHKXMMLDIKGLELPIGNP-----AMSW 924

RESULT 10
Q92P04
ID Q92P04 PRELIMINARY; PRT; 935 AA.
AC Q92P04;
DT 01-MAY-1999 (TremBLrel. 10, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE Alpha-D-xyllosidase precursor.
OS Tropaeolum majus (Common nasturtium).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Tropaeolaceae; Tropaeolum.
OX NCBI_TaxID=4020;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. tall Climbing mixed; TISSUE=Cotyledon;
RA Crombie H.J., Chengappa S., Jarman C., Sidebottom C., Reid G.;
RT "Molecular characterisation of a xyloglucan oligosaccharide-acting
RT alpha-D-xyllosidase from the cotyledons of nasturtium (Tropaeolum majus
RT L) seeds.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ131520; CAA10382.2; -.
DR InterPro: IPR000322; Glyco_hydro_31.
DR Pfam: PF01055; Glyco_hydro_31; 1.
DR PROSITE: PS00129; GLYCOSYL_HYDROL_F31_1; 1.
KW SIGNAL.
FT CHAIN
FT SIGNAL
SQ SEQUENCE 935 AA; 104937 MW; 22DE6901E9C19BD CRC64;

Query Match 42.1%; Score 1950.5; DB 10; Length 935;
Best Local Similarity 45.6%; Pred. No. 5.2e-136;
Matches 401; Conservative 134; Mismatches 263; Indels 81; Gaps 16;

Qy 59 DVORLAVASLTGRLRVRITDADHPWEVPODIIIPAPGDVLHDAPPASSAPLOGR- 117
Db 74 DIPILQLYVHESQRLRVRITDADHPWEVPODIIIPAPGDVLHDAPPASSAPLOGR- 128
Qy 118 ---VLSPAGSDLVLTWH-ASPRFTVSRSTGDTLFDTPAG-----LVFRDKYLEVTS 166
Db 129 LITVSEISGSELIFSYPDRDRFGFAVKRKSNGETLNFSSDPSDPFGEMVFKQVLEIST 188
Qy 167 ALPAGRASLYGGEHTK-SSFLRLHNDSTLWNADIGASYVDVNLVYSGHPFMDVRAPG- 224
Db 189 KLPKD-ASYIGUGENTQPHGILKYFNDPYLLTMDVSNALNADSYGSHPMYDLNRVGG 247
Qy 225 --TAHGVLSSNGMDVLYGGSYTVYKVIIGVLDYFFAGNPLAVDQYQTQIARPAPM 282
Db 248 EAYAAVLLNSNGMDVYFGDLSLYKIIGVDFVFFIGPAPLDVQYQYTAIGRPAPM 307
Qy 283 PWSFGHCRGYLVNSDLERVARYAKARPLEVMTDDIDYMGFKDPTLDRVNFATA 342
Db 308 PYWSLGFHCQWGYHNLVIEDVVERVYKNAKIPLDVIMNDDHMDGKDKDTLNTKNYPRP 367
Qy 343 ELRPFFVRLHRAQKVLILDPGIRVDPIDATYGVTFVRCMQDIDFLKNGTNFVGNWPG 402
Db 368 QLLAFLDKXISGIGMKYIIVDPIGIAV---NSSGYTQRGLANDVFIKTEGEPFLAQWPG 424
Qy 403 DVYFPDFMPHAAAEFAWEISLFRKTPVDGLWIDMNEISNF-----YNPE----- 448
Db 425 AVNFPDFLPKPTVDMWGDVRRFHELVPVDGLWIDMNEISLFLFWMENPQCKQCTGEG 484

```

```

Qy 449 -----PMNALDDPPYRINNDCTGRPINNKTYRPLAVHYGGVTEYEHNLFGLL 496
Db 485 PGWICCLCKNITKTRDEPPYKINATGVQAPIGFKTIATSTCTHYNGVLEYDAHSIYGF 544
Qy 497 EARATGRGLRDTGRPPVLSRSTFVSGSGRYTAYMTGDNAATWGDRLRYSTINTLSFGLPG 556
Db 545 QTVAHKALQGLEGKRPILTRSTFVSGSGHYAAHWTDGNOGTWENLRYSTINTLSFGLPG 604
Qy 557 MPNIGADICGNGNTTEELCGRWIOLGAFYFSDKHSIAFTVRRELYLWPSVAASGRKAL 616
Db 605 VPMVGSIDCGFYPOPTBELCNRWIEVGAFYFSDKHSIAFTVRRELYLWPSVAASGRKAL 664
Qy 617 GLRYOLLFPVYLYEAAHMTGAPARPLFFSYPHDVATYGVDRFLLCRGVLSVPVLEPG 676
Db 665 GMYKLLPFLVLYEAAHMTGAPARPLFFSYPHDVATYGVDRFLLCRGVLSVPVLEPG 724
Qy 677 PTVDAYFPAGRWYRLDYSLAVATRTGKHVRLPAPADTVNVHLTGGTILPQGSALTTS 736
Db 725 KTEVKALPPGWTWYSLDWTETVDSKQGVVTLAPLHVNVNHLVQNTILPMQOGLLSK 784
Qy 737 RARATAFHLLV---ALAEDGTASGVFLDGDGSPYGRSDMS-MVRFNYPKPNKGAIK 792
Db 785 EARTMFTLIVTPPAGATDQKAGNLFLLDKDELPMKNGISYTYVEFYATL--NQGAVK 842
Qy 793 VKSEVHNSYAQSRTLVIVSKVLMGHRSPAPKULTVHNSAEVASSAG-----T 844
Db 843 VMSQVQEGKFDALDNGWSIEKVTVLGSNK-----QVGSLEIDGSPVSGISKVEMSS 894
Qy 845 RYQNAAGLGG-----VAHIGGLSLVVGEEFELKVAM 875
Db 895 EQIFVQKGLDAESKPESLMVEVKGDLDFVGNKFNVMWOM 933

RESULT 11
Q9LZT7
ID Q9LZT7 PRELIMINARY; PRT; 855 AA.
AC Q9LZT7;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Hypothetical 95.9 kDa protein.
GN F16L2.150.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Jordan N., Bangert S., Wiedelmann R., Voss H., Unseld M., Mewes H.W.,
RA Rudd S., Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL162459; CAB82818.1; -.
DR InterPro: IPR000322; Glyco_hydro_31.
DR Pfam: PF01055; Glyco_hydro_31; 1.
DR PROSITE: PS00129; GLYCOSYL_HYDROL_F31_1; 1.
KW Hypothetical protein.
SQ SEQUENCE 855 AA; 95858 MW; 54CB7A835BB1C50E CRC64;

Query Match 41.1%; Score 1904; DB 10; Length 855;
Best Local Similarity 43.8%; Pred. No. 1.5e-134;
Matches 404; Conservative 144; Mismatches 245; Indels 130; Gaps 21;

Qy 1 MATVGVLLLCCLCLCLFAPRLCS-----SKEGEPLAARTVLAVVTMEGALRAEA 50
Db 1 MASCLSLVAILLC-FSSIQCSNAIGKGYRLISMESKSPDGSFI-----GYLQVK 51
Qy 51 TGRSRSTGDVQRLAVVAYSLTDSRLRVRITDADHPWEVPODIIIPAPGDVLHDAPPAS 110

```



```

Db 720 TVARPLFEFFWDVATYGLDQFLWGLSLLVTPVLEFCADSVLGYFQGVWYDPTGS-- 777
Qy 699 VATRTGKHVRILPAPADTVNVHLTGTTGLPLQOQSALTTSRRRTAFHLLVALAEDGTASGY 758
Db 778 SVNSSGEMKLKSLAPLDHLNLHLREGSLPTQKPGITSKATGNPLHLILVALSTRATWGD 837
Qy 759 LFLDGDSPGPEGRSDASWVFNFKIPNNKCAIK--VKSEVHNSYAQSRTLIVSKVLM 816
Db 838 LFWDDGESLDTFEOGNSYLVFN-----ATENIFTSNVLHAS-TEATDVTDIVASFY 898
Qy 817 GHRSPAAPKLLTVHNSAEVASSAGTRYQVAGGLGVGAHIGGLSLVVGEEFELK 872
Db 889 GVQEP--PSKVL-----LDQEKPSYLDNQ-----VLTVSLGLVLSQGSFLQ 930

RESULT 13
Q9UV08
ID Q9UV08 PRELIMINARY; PRT; 992 AA.
AC Q9UV08;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Alpha-glucosidase Agda.
GN AGDA.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Peiziomycotina; Eurotiomycetes;
OC Eutotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=162425;
RN [1]
RP SEQUENCE FROM N.A.
RA Boase N.A., Murphy R.M., Kelly J.M.;
RT "An amylose cluster in Aspergillus nidulans.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF208225; AAF17102.1; -.
DR InterPro; IPR001064; Crystallin.
DR InterPro; IPR00322; Glyco hydro 31.
DR Pfam; PF01055; Glyco hydro 31; 1.
DR PROSITE; PS00225; CRYSTALLIN BETAGAMMA; 1.
DR PROSITE; PS00129; GLYCOSYL HYDROL_P31_1; 1.
DR PROSITE; PS00707; GLYCOSYL-HYDROL_F31_2; 1.
SQ SEQUENCE 992 AA; 109666 MW; 30FB3B887345A72E CRC64;

Query Match 31.2%; Score 1443; DB 3; Length 992;
Best Local Similarity 35.6%; Pred No. 1,le-99;
Matches 330; Conservative 159; Mismatches 283; Indels 156; Gaps 27;

Qy 59 DVQRLAVASLETGRLVRIT----DADHPRWEV-PQDIIIPRAPGDLVLDHAPPASSAP 113
Db 106 DVESLTLMQYQDTRNLQITPTTYVDASNASWYLPPEFVRP-----KPAAGA- 155
Qy 114 LQGRVLSPAGSDLVLTVAHP--FRFTVSRRTGDTLPTAPG-LVFRDKYLEVTSAIPAG 171
Db 156 -----SESHDFATWNEPTFNQVTKSTGEVLFDTAGSVLFENQFIEFTVSLPE- 208
Qy 172 RASLYGLGEHTKSSFLRHNSFTLWNADIGASYVDVNLVYGHSHFFYMDVR----- 221
Db 209 EYNLYGLGERI-NQLRLRNATLTSYAADIG-NPIDANIYGHAFYVDTRYFVSDEAGKH 266
Qy 222 -----APGTAGVLLLSNGMDVLVYGGSVYTKVGLVDFFYFAGPNPLAV 268
Db 267 TVYKSSADPSATYTSYSHGVFLRNSHGEVVLNPQGLTWRTIGSDILTYSGFTVAEV 326
Qy 269 VDQYTO-LIARPAIPYNSFGHCQRVGLVNVSLDERVAVAKARIPLEVMWTDIDYMD 327
Db 327 TKQYQSTVGLPAMQKYDTLGHQCRWGNWNSVADVLANPEKFEIPLEVLKADIDMH 386
Qy 328 GKDFTLDRVNFATAELRPFVDRLRHNAQKYLILDPGIRV-DPIDA--TYGTFVRGMQ 384
Db 387 GYRNFENDEYRPPYNETKVLFDKLHAGGRHFVPIVDAALYIPNPNQNASDSYETVTRGAAR 446
Qy 385 DFLKLR-NGTNFVGNVWPGDVFFPDMHPAAAEFFAREISLFRRTIPVDGLWIDWNEISN 443

```

```

Db 447 DVFLKXPDGSLYIGAWPGYTVFPDHHHPDAADFWANELVTVYKVKFDGWWYDMSEVSS 506
Qy 444 F-----YNP-----EPMNA----- 452
Db 507 FCVSGSGSRNRLNPNVHPPLPGFPGVGNVDYVEYPGFELSNAEASASASSQAATTA 566
Qy 453 -----LDDDPYBINNDGTCGRPINNKTYRPLAVHYGVGVTEYEHN 491
Db 567 TTTTSTSYLRTSTPTGVRNVNYPYVINHVQTHDHLAVHAVSPNATHVDGYHYDVHS 626
Qy 492 LFGLEEARATGRVLRDTCGR-RPFVLSRSTFVSGRYTAYMTGDNAATWGLDRLRYINTML 550
Db 627 LYGHMGIQATYRGLTQIAPRKEPFIIGSTFAGSKWAGHMGDNYSRWSSMYFISQAL 686
Qy 551 SFGLFQMPMIGADICGFNGNTTEELCGRWIQLGAFYFSPSRHSAIPTVRRELYLWPSVAA 610
Db 687 QFSLYGIPMGFVDTGCGFSGNTABEELCNRMQLSAFFFPYRNHNVLGTIPQEPYQWASVID 746
Qy 611 SGRKALGLRYQLLPYFYLTMVEAHMTGAPRLPFFSYPHDVATYGVDRQFLGLGRVLVS 670
Db 747 ATKKAWRIYALLPYFYLMDHDAHTTGSTVLBALAWBFPDDSLAIDNQFLVGSILVT 806
Qy 671 PVLEPGPTTVDAYFP-AGR---WYSLYDYSLAVATRTGKHLVPAPADTVNVHLFGGITL 726
Db 807 PVLEPGVSTVKGVFGVGGQGEVWYDWTQT-AVDAQPGVNTTIDAPLGHIPVYVRGGSIL 865
Qy 727 PLOQSALTTSRRARTAPHLIVALAEDGTASGLVFLDDGDS--PEYGRSRDSMSWVFNKYI 784
Db 866 PMQEPALTTRDARKTPWALLVALGKDGTSAGHLYLDDGESIHPK-----VSLNVPK 916
Qy 785 PNNKGAIKVKSEVHNSYAQSRTLIVSKVLMGHRSPAAPKLLTVHNSAEVASSAGT 844
Db 917 RATQFALTYSSE---GEWKEANPL--ANVTILG-----VLENPVSVTSNGQVPAEYDAQS 967
Qy 845 RYONAGGLGVGAHIGGLSLVVGEEFELK 872
Db 968 RILVITGLNQFTNGAW-----QQDWTLR 991

RESULT 14
Q8BG16
ID Q8BG16 PRELIMINARY; PRT; 953 AA.
AC Q8BG16;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Glucosidase.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J, and NOD; TISSUE=Heart, and Thymus;
RX MEDLINE=22354683; PubMed=1246851;
RA The PANTOM Consortium.
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK052211; BAC34888.1; -.
DR EMBL; AK088481; BAC40382.1; -.
SQ SEQUENCE 953 AA; 106247 MW; 956B89685FB5FF81 CRC64;

Query Match 31.1%; Score 1439; DB 11; Length 953;
Best Local Similarity 38.6%; Pred. No. 2e-99;
Matches 331; Conservative 124; Mismatches 317; Indels 86; Gaps 20;

Qy 14 CLFAPRLCSSKEGPLAARTVLAATVMEGALRAEAATGGRSSTGVDQRLAVASLETDS 73
Db 127 CFPSPSPYPSYRLN-----LSTESGYTATLTRSPFTFPKDVLTQLEVLMTSDS 177
Qy 74 RLVRITDADHPRWEVPODIIIPRAPGDLVLDHAPPASSAPLQGRVLSPAGSDLVLTVAHS 133

```

```
Db 178 RLHFKIDPASKRYEVPLE-TPR-----VLSQAP-----SPLYSVFES-----EE 216
QY 134 PFRFTVSRRTGTDLFT--APGLVFRDKYLVTSALPAGRASLYGLGHEHTKSSFFLRHN 191
Db 217 PFGVVRKLGGRVLLNTTVP-LFFADQFLQLSTLPS--QHITGLGEHLSPLMLSTDW 273
QY 192 DSFTLWADIGASVVDVNLVYSGHPFYMDVRAPGTAHGVLLSSNGMDVLYGGS-VYTVKV 250
Db 274 ARITLWNRDTPPSQ-GTNLYGSHPFYLALEDGGLAGVFLNLSNAMDVLQSPALTWRS 332
QY 251 IGVLDVFFPAGNPLAVDQYQTOLIAARPAMPYSGFHQCRGYLNVSDLERVARYA 310
Db 333 TGGILDVYVFLGPPEKSVQOYLVDVGYPPMPYVGLFCRWGYSSTAIVRQVVENMT 392
QY 311 KARPLEVMWTDIDYMDGFKDFTLDRVNFVTAELRPFVDRLHRNAKYVILDPGIRVDP 370
Db 393 RTHFPDLVQNDLDYMDARDFTFNQDSF--ADFPDMVRELHQGRRYMIVDPAISSAG 450
QY 371 IDATYGTFRGMOQDIFL-KRNGTNFVGNVMPGDVVPDPMHAAAEFWAREISLFRTI 429
Db 451 PAGESPYDEGLRGVFTITNETGOPLIGKVPWGTAFPDFTNPETLDWQDMVSEFAQV 510
QY 430 PVDGLWIDMNEISFNYPE-----PMALDDPPYRINNDGTRPINNKTVRPLAVHYGGV 485
Db 511 PFDGMMLDMNEPSNFRVSGSQGCPNNELENPY---VPGVVGILQQAATICASSHQFLST 567
QY 486 EYEBHNLFGLLEARATGRGVLRDTRGRPFVLSRSTFVSGSRYTAYWTGDNAAATWGLDLYS 545
Db 568 HYNLHNLGLTEAIASSRALVKTRGTRPFVLSRSTFSGHGRYAGHTGDRSSWEHLAYS 627
QY 546 INTWLSFGLFPMGIGADI CGFNCTTEELCGRIQI GAFYPPFSRDSHAIFTVRRELYW 605
Db 628 VPDILQFNLLGVPLVGADICGFIGDTSEELCVRWTLQAGFYPPFMRNNDLNSVPQBPYRF 687
QY 606 PSVAASG-RKALGLRYOLLPYFTLWYEAHMTGAPIARPLFESYPHDVATYGVDRQFLG 664
Db 688 SETAQAMRKAFALRYALLPYLYTLFRAHVRGDTVARPLFLEFPEDPSTWSDRQLLWG 747
QY 665 RGLVSPVLEPGPTTVDAYFPAGRWYRLYDYL-----AVATRTGKHV 707
Db 748 PALLITPVLEPGKTEVTGYFPKGTWYNNQMVSDSLGTLPSPSASSFRSAVQSK-QQWL 806
QY 708 RLPAPADTVNHLTGTTILPQOQSALTTSRARTAFHLLVALAEDGTASGYLFDGDSGP 767
Db 807 TLEAPLDTINVHLREGYIIPLOQSULTTTSRQPMALAVALTASGEADGELFWDDGESL 866
QY 768 EYGRSDWSMVRNYKIPNN-----KGAIKVKEVHNYSYAQSRTLVIKSVLM 816
Db 867 AVLERGAYTLVTFSAK--NNTIVNKLVRVTKEGAEQLQREVTVLG VATAPTQVLSNGIPV 924
QY 817 G-----HRSPAAPKKL 827
925 SNETYSPDNKSLAIPVSL 942
```

## RESULT 15

```
Q91245 ID Q91245 PRELIMINARY; PRT; 953 AA.
AC Q91245;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Similar to glucosidase, alpha, acid.
GN GAA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N. A.
RA Straubeberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
```

```
DR EMBL; BC010210; AAH10210.1; --
DR MGD; MGI:95609; Gaa.
DR InterPro; IPR000322; Glyco_hydro_31.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR SMART; SM00018; trefoil; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
DR PROSITE; PS00025; P_TREFOIL; 1.
SQ SEQUENCE 953 AA; 106157 MW; 320B6D7354A5FF91 CRC64;

Query Match 31.1%; Score 1438; DB 11; Length 953;
Best Local Similarity 38.6%; Pred. No. 2.4e-99;
Matches 331; Conservative 123; Mismatches 318; Indels 86; Gaps 20;

QY 14 CLFAPRLCSCKEEPLAARTVLA VAVTMEGALRAAAATGCRSSTGDVORLAVYASLTDS 73
Db 127 CFFPPSPVSYRLN-----LSSTESGYTATLRTSTPTFFPKDVLTLQLVLEMETDS 177
QY 74 RLVRITDADHPREVPODIIIPRAPGDVLDHAPASSAPLOGRVLS PAGESDLVITVHAS 133
Db 178 RLHFKIDPASKRYEVPLE-TPR-----VLSQAP-----SPLYSVFES-----EE 216
QY 134 PFRFTVSRRTGTDLFT--APGLVFRDKYLVTSALPAGRASLYGLGHEHTKSSFFLRHN 191
Db 217 PFGVVRKLGGRVLLNTTVP-LFFADQFLQLSTLPS--QHITGLGEHLSPLMLSTDW 273
QY 192 DSFTLWADIGASVVDVNLVYSGHPFYMDVRAPGTAHGVLLSSNGMDVLYGGS-VYTVKV 250
Db 274 ARITLWNRDTPPSQ-GTNLYGSHPFYLALEDGGLAGVFLNLSNAMDVLQSPALTWRS 332
QY 251 IGVLDVFFPAGNPLAVDQYQTOLIAARPAMPYSGFHQCRGYLNVSDLERVARYA 310
Db 333 TGGILDVYVFLGPPEKSVQOYLVDVGYPPMPYVGLFCRWGYSSTAIVRQVVENMT 392
QY 311 KARPLEVMWTDIDYMDGFKDFTLDRVNFVTAELRPFVDRLHRNAKYVILDPGIRVDP 370
Db 393 RTHFPDLVQNDLDYMDARDFTFNQDSF--ADFPDMVRELHQGRRYMIVDPAISSAG 450
QY 371 IDATYGTFRGMOQDIFL-KRNGTNFVGNVMPGDVVPDPMHAAAEFWAREISLFRTI 429
Db 451 PAGESPYDEGLRGVFTITNETGOPLIGKVPWGTAFPDFTNPETLDWQDMVSEFAQV 510
QY 430 PVDGLWIDMNEISFNYPE-----PMALDDPPYRINNDGTRPINNKTVRPLAVHYGGV 485
Db 511 PFDGMMLDMNEPSNFRVSGSQGCPNNELENPY---VPGVVGILQQAATICASSHQFLST 567
QY 486 EYEBHNLFGLLEARATGRGVLRDTRGRPFVLSRSTFVSGSRYTAYWTGDNAAATWGLDLYS 545
Db 568 HYNLHNLGLTEAIASSRALVKTRGTRPFVLSRSTFSGHGRYAGHTGDRSSWEHLAYS 627
QY 546 INTWLSFGLFPMGIGADI CGFNCTTEELCGRIQI GAFYPPFSRDSHAIFTVRRELYW 605
Db 628 VPDILQFNLLGVPLVGADICGFIGDTSEELCVRWTLQAGFYPPFMRNNDLNSVPQBPYRF 687
QY 606 PSVAASG-RKALGLRYOLLPYFTLWYEAHMTGAPIARPLFESYPHDVATYGVDRQFLG 664
Db 688 SETAQAMRKAFALRYALLPYLYTLFRAHVRGDTVARPLFLEFPEDPSTWSDRQLLWG 747
QY 665 RGLVSPVLEPGPTTVDAYFPAGRWYRLYDYL-----AVATRTGKHV 707
Db 748 PALLITPVLEPGKTEVTGYFPKGTWYNNQMVSDSLGTLPSPSASSFRSAVQSK-QQWL 806
QY 708 RLPAPADTVNHLTGTTILPQOQSALTTSRARTAFHLLVALAEDGTASGYLFDGDSGP 767
Db 807 TLEAPLDTINVHLREGYIIPLOQSULTTTSRQPMALAVALTASGEADGELFWDDGESL 866
QY 768 EYGRSDWSMVRNYKIPNN-----KGAIKVKEVHNYSYAQSRTLVIKSVLM 816
Db 867 AVLERGAYTLVTFSAK--NNTIVNKLVRVTKEGAEQLQREVTVLG VATAPTQVLSNGIPV 924
QY 817 G-----HRSPAAPKKL 827
```

Db 925 SNFYSPDNKSLAIPVSL 942  
:|:|:|

Search completed: October 27, 2003, 10:30:52  
Job time : 69.995 secs







```

Db 239 YGSHFFYMDVRSFVAGSTGVHLLNSNGMDVEYTGNRITKYKVIIGGIDLYFFAGFSPGQ 238
QY 268 VDOQYTOIARPAAMPYMSFGHOCRYGLYNVSDLERVAVYAKARIPLVEMWTDIDYMD 327
Db 299 VVQFTRVIGRPAAMPYAFWFGQCRGYHDVVELQSVAGYAKARIPLVEMWTDIDYMD 358
QY 328 GFKDFTLDRVNTAAELRPVFDRLHRAQAQVILIDPGRVDPIDATYGTFFVRGMOODIF 387
Db 359 AYKDFTLDPVNFPLDKMKKFVNNLHKNGQKYVILDPGIST---NKTYYEYIRGMKHDPV 415
QY 388 LKRNQTNFVGNVPGDYVFPDFMHPAABFWABISLFRPTIPVDGLWIDMNEISFYN - 446
Db 416 LKRNKPYLSVMPGPYVFPDFLKSALTFTWDEIKRFLNLLPVDGLWIDMNEISFISS 475
QY 447 -PEPMALDDPPYRINNDGTRPINNKTVRPLAVHYGGVTEYEEHNLFGLEARATGRGV 505
Db 476 PPIPGSTLDNPPYKINNSGVMLPIINTIPTAMHYGDIPEYVNNHLFGYLEARVTRAL 535
QY 506 LRDTRRPFVLSRSTFVSGRYTAYWTGDNAATWGLDRLYSINTMLSGFLGPMPIGADIC 565
Db 536 IKLTEKRPVLSRSTFSGSKYTAHWTGDNAATWDLVYSIPSNLDFGLFPIPMVGADIC 595
QY 566 GFNGNTEELCGMIOLGAFYPSRDHSAFTVRELYLWPSVAASGRKALGIRYOLLPY 625
Db 596 GFLGNTTEELCRMIQAGAFYPSRDHSSLGTYQELRYWESVAASARKVYLRYTLPPY 655
QY 626 FYTLMYEAHMTGAPRIARLPFFSYPHDVATYGVDRQFLGLGVLSVPLEPGPTTVDAYPP 685
Db 656 FYTLMYEAQLNGPIARLPFFSPDDIKTYGISQFLLGKGMVSPVLKGVSVVTAYPP 715
QY 686 AGRWRYLYSVLAVATRGKHVRLPAPADTVNHLTGTTILPQQSALTTSRRARTAFHL 745
Db 716 RGNWFDLFDVRSVTASTGRVYVTLASPDHINHVIQENILAMQGMKAMTTOAARKTFPHL 775
QY 746 LVALAEDGTASGLYFLDDGDSPEVG--RSDWSVRFNKYIPNNKGAIKVSEVHNSYAO 804
Db 776 LVMSDCGASGFLGLDDGVEYTWGNVRGKWTYKFE--IAASAKQTCIITSDVVSGEFV 833
QY 805 SRTLIVSKVLMGHRSPAAPKPLTVHVNSEAIVEASSAGTRYONAGGLGVGAHIGGLSLV 864
Db 834 SQKWIDKVTILGURKTKINGYVTRGATVRKGDASKLSTPDRKGEFIVAEISGLNLL 893
QY 865 VGBEFEL 871
Db 894 LGREFKL 900

RESULT 3
AGLU BETVU STANDARD; PRT; 913 AA.
AC O04931;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase) .
OS Beta vulgaris (Sugar beet) .
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Caryophyllaceae; Chenopodiaceae; Beta.
OC NCBI_TaxID=161934;
RN [1]
RP STRAIN=CV. NK-152;
RC MEDLINE=97321863; PubMed=9178565;
RA Matsui H., Iwanami S., Ito H., Mori H., Honma M., Chiba S.;
RT "Cloning and sequencing of a cDNA encoding alpha-glucosidase from
RL sugar beet.";
RL Biosci. Biotechnol. Biochem. 61:875-880(1997) .
RN [2]
RP ACTIVE SITE, AND SEQUENCE OF 464-472.
RX MEDLINE=95252592; PubMed=7766184;
RA Iwanami S., Matsui H., Kimura A., Ito H., Mori H., Honma M.,
RA Chiba S.;

```

```

RT "Chemical modification and amino acid sequence of active site in
RT sugar beet alpha-glucosidase.";
RL Biosci. Biotechnol. Biochem. 59:459-463(1995) .
CC -!- FUNCTION: HIGH ACTIVITY FOR ALPHA-GLUCAN.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-
CC linked D-glucose residues with release of D-glucose.
CC -!- PTM: THE N-TERMINUS IS BLOCKED.
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D89615; BAA20343.1; -.
DR PIR; JCS463; JCS463.
DR InterPro; IPR00322; Glyco hydro_31.
DR Pfam; PF01055; Glyco hydro_31.1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
KW Hydrolase; Glycosidase; Glycoprotein; Signal.
FT SIGNAL 1 28
FT CHAIN 29 913
FT ACT_SITE 469 469
FT CARBOHYD 54 54
FT CARBOHYD 404 404
FT CARBOHYD 495 495
FT CARBOHYD 517 517
FT CARBOHYD 728 728
FT CARBOHYD 823 823
FT SEQUENCE 913 AA; 102117 MW; 02AC4F08505369CC CRC64;
Query Match 51.2%; Score 2370; DB 1; Length 913;
Best Local Similarity 51.2%; Pred. No. 1.1e-162;
Matches 466; Conservative 147; Mismatches 235; Indels 62; Gaps 15;
QY 7 LLLCLICLFLPRLCSSKEEGPLAARTVLAVATMEGALRAEAATGGRSSTG----- 58
Db 16 VVPLVLCMVVEGATSKNDNQGEA---IGYQYQKNA-KVDNST-GKSLTALLQLRNS 70
QY 59 -----DVQRVAVYASLETDSRLRVITDADHPWVQDIIIPRAPGVLDHAPASSAP 113
Db 71 PVYGPDIHFLSTASFEEDTLRIFTDANRWWEIPNEVLRP-----PPPSPPP 122
QY 114 LQG-----RVLPAGSDLVLTV-HASPFRTVSRSTGDTLFDATP----- 153
Db 123 LSSLQHLKPIPNQPTTTLVLSHPSDLAFTLFTHTTPTGFTIYRKSTHDLVFDATPIPSN 182
QY 154 ---GLVFRDKYLEVTSALPAGBASLYGLGHTKSSFLRHNDSPFLMNADICASYDVNL 210
Db 183 PTFFLYIKDYLQLSSSLPAQQAHLVGLGHTKPTFQLAHQNLILLMNADIASFNRLNL 242
QY 211 YGSHPFYMDVRAAP---GTAHGVLLSSNGMDVLYGGSVTVYKVGVDLDFPFAGNPILA 267
Db 243 YGSHPFYMDVRSPPMVGTHGVFLNSNGMDVYVYTGDRITKYKVIIGIDLYFACTREM 302
QY 268 VVDQYTOIARPAAMPYMSFGHOCRYGLYNVSDLERVAVYAKARIPLVEMWTDIDYMD 327
Db 303 VLDQYTKLIGRAPMPYNAFPHQCRWGYRDVNEIETVVDKYABARIPLVEMWTDIDYMD 362
QY 328 GFKDFTLDRVNTAAELRPVFDRLHRAQAQVILIDPGRVDPIDATYGTFFVRGMOODIF 387
Db 363 AFKDFTLDPVNFPLDKMQQFVTKLHNGQRYVPIIDPGINT---NKS YGTFFIRGQSNVF 419
QY 388 LKRNQTNFVGNVPGDYVFPDFMHPAABFWABISLFRPTIPVDGLWIDMNEISFYN - 446
Db 420 IKRNQNPYLGSMVPGVYVTPDFLDPAAKSFWDKIKRFDIPLFDGIWIDMNEASNFITS 479
QY 447 -PEPMALDDPPYRINNDGTRPINNKTVRPLAVHYGGVTEYEEHNLFGLEARATGRGV 505

```

```
Db 480 APTGSLDNPYPKINNSGGVRPINSKTIPTATAMHYGNVTEYNAHNLGYFLESOATREAL 539
Qy 506 LROTRPFPVLSRTFVSGRYTAIWTGDNAATWGLDRLYSINTMLSPGLGMPMIGADIC 565
Db 540 VRPATRGPFLLSRTFAGSGKYTAHMTGDNAARWDDLQYSIPTMLNFGLEGMPMIGADIC 599
Qy 566 GFNGTTEELCGRIOLGAFVPPSRDHSALFTVRELYLWPSVAASGRKALGLRYQLLPY 625
Db 600 GFASSTTEELCCRWIQLGAFVPPSRDHSADTTQOELYLWESVAASARTVGLRYELLPY 659
Qy 626 FYTLMYEAHMTGAPARLPFFSYPHDVATVQDROFLLRGVLVSPVLEPGPTTVDAYFP 685
Db 660 YITLMYDANLRGSPARLPSTFPDDVATVGISSQFLIGRMVSPVLPQSSIVNAYS 719
Qy 686 AGRVRLYDYSLATVTRTKVRLPAPADTVNHLTGTTLPLQOASALTTSRARTAFHL 745
Db 720 RGNVLSLNTSSVSAGTVLSAPPDHIHNEGINVAMQCEAMITTAARSTPPHL 779
Qy 746 LVALAEDGTASGYLFLDDGDSPEY-GRSDMSVMRFNYKIPNNKGAIKVSEVHNSYAQ 804
Db 780 LVMSDSHVASTGELFLDNGIEMDGGPGKWTLVRFPAESGINN--LTISEVVNRGYAM 837
Qy 805 SRTLVISKVLMGHRSPAAPKXLTHVNSAEVBEASS-SAGTRYQNAAGLGGVAHIGGLSL 863
Db 838 SQRWMDKITILGLKRVKIKYTVQKDAGAIKVKGLGRRTSSHNQGGF-FVSVISDLRQ 896
Qy 864 VVGEEPELVK 873
Db 897 LVGQAFKLEL 906

RESULT 4
AGLU_MUCJA STANDARD; PRT; 864 AA.
ID Q92442;
AC 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase).
OS Mucor javanicus
OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
OC Mucor.
OX NCBI_TaxID=51122;
RN STRAIN=IFO 4570;
RC STRAIN=IFO 4570; PubMed=8830045;
RX MEDLINE=96271012;
RA Sugimoto M., Suzuki Y.;
RT "Molecular cloning, sequencing, and expression of a cDNA encoding
  alpha-glucosidase from Mucor javanicus.";
RL J. Biochem. 119:500-505(1996).
CC -!- FUNCTION: Hydrolyzes not only malto-oligosaccharides but also
  soluble starch.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-
  linked D-glucose residues with release of D-glucose.
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D67034; BAAL1053.1; -
CC InterPro; IPR000322; Glyco_hydro_31.
CC Pfam; PF01055; Glyco_hydro_31; 1.
CC PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
CC PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; FALSE_NEG.
CC Hydrolase; Glycosidase; Glycoprotein; Signal.
CC SIGNAL 1 22 POTENTIAL.
CC CHAIN 23 864 ALPHA-GLUCOSIDASE.
```

```
FT ACT SITE 430 430 BY SIMILARITY.
FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 364 364 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 406 406 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 500 500 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 734 734 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 864 AA; 98761 MW; CFAB4759DC431403 CRC64;

Query Match 31.7%; Score 1470; DB 1; Length 864;
Best Local Similarity 36.9%; Pred. No. 6.3e-98;
Matches 328; Conservative 158; Mismatches 281; Indels 122; Gaps 25;

Qy 1 MATVGVLICLCLCFAPRLCSSKEEGPLAATVLAVALT-----MEGALRAEAA--- 50
Db 1 MAKVSFIFVATALI-----TGNVLCQTATAYAVSSSAFGYKIDGHVRKTEAGLH 49

Qy 51 -----TGRSGSTG-----DVORLAVVASLETDSRLRVIRITDADHPRWEVPQDIIPRA 98
Db 50 IPLTLNSRGNKKTGIDTFGKTIKDITVDVYVETEERLHVKISDKAKQYLYP----- 101

Qy 99 PGDVLHDAPPASSAPLOGRVLSPAGSDLVLTVHASPPFTVSRRTSGDTLFDTA-PGLVLF 157
Db 102 -----DSPLOGERPOIKHYVSPKHSNLDFOVTAKPFSKVVVRKDDKTTIFDTNMLPVF 155

Qy 158 RDKYLEVTSALPAGRASLYGLGEHTKSSFLRHNDSPFLMNADIGASVVDVNLGSHSPY 217
Db 156 EDQYLELSTKVPED-ANIYGIGEVY-APFRTHNVT-TLWARDNPDFFYR-NIYGAHPY 211

Qy 218 MDVRAPGTAGHGVLLSSNGMDVLYGGSYTVKYVIGGVLDYFFFA----GPNPLAVVDQYT 273
Db 212 QEVYR-DGKAHGALLANAHGMDVITTEGRITKYVIGGILDFYFFAPKSGKPNDSIA--YT 268

Qy 274 QLIARPAWPMYWSFGHOCRYGYLNVSDLERVARYAKARPLEVMWTDIDYMGDFKOPT 333
Db 269 DLIGKPMPSHMLGMWCHCRYGPNIDKVTYVGRKYKEANIPLOTVMYDIDYMEETKOPT 328

Qy 334 LDRVNFATAELRPFVDRLHRNAQKVLLIDPGIRVDPIDATYGTFRVGMQODIFLKR-NG 392
Db 329 FKVNFPPQDRMIGLGEQLHKQGNVVMVDPAISA---NTTYEPYVRGTENDVMTKNADG 385

Qy 393 TNFVGNVMPGDVYFPDFMHPAAAEFWAREISLFRRTIPVDGLWIDMNEISNF----- 444
Db 386 SDFIGSVMPGFTTFPDWWHPNATKYWNKEIIDFVDMLGVDGLWIDMNEPASFCLGSCGSG 445

Qy 445 -----YNPE-----PMNALDDPP-----YRINDGTGRPINN 471
Db 446 KYDAGNQPYRWTYTEEQAAHNRWEKSLKAMGNPPGEERNLLYPKYAINN-GAGN-USE 503

Qy 472 KTVRPLAVHYGVTEYEENHFLGLEABATGRGVLRDTR-RPFVLSRSTFVSGRYTAY 530
Db 504 FTVATTALHYGNI PHYDIHNLGHAESHITQALIKHKNKIRPFVLTSSFPGSKSVGH 563

Qy 531 WTGDNAATWGLDRLYSINTMLSPGLGMPMIGADICGFNGNTTEELCGRIOLGAPYPSR 590
Db 564 WTGDNHSFPYPLKNGSIANILNFQMFVSGYSGADVGCFNSDTTEELCTRMEIGAFYFAR 623

Qy 591 DHSALFTVRELYLWPSVAASGRKALGLRYQLLPYFYLTMVEAHMTGAPARLPFFSPH 650
Db 624 NNNNNNAKQOEPYLMWESTAESRIANTRYEMLPFYTLFEESNRLGLGWRLPFLFEPA 683

Qy 651 DVATYGVDRQFLLRGVLVSPVLEPGPTTVDAYFPAGRWRLYLDYSLAVATRTGKHYR-- 708
Db 684 YEELYSNDVQTLVSGDILLSPVLDSEKTSVKAQFPGGQWYDWTYHETLVNKSNNKVKTV 743

Qy 709 -LPAPADTVNHLTGTTLPLQOASALTTSRARTAFHLVALAEDGTASGYLFLDDGSP 767
Db 744 TLDAPLTHIPHIRGCAIIPTKYTYTGETPATPYNLVIALDKKGQASGRLYIDDGESL 803

Qy 768 EYGRSS-----DMSMVRFNKYIPNNKGAIKVSEVHNSVAQSRTLV 809
Db 804 EVKSSGVHFHLQBSWSPGF-QWVWLQBG-----RKDWLHHHTWQARQV 847
```

## RESULT 5

AGLU\_ASPNG STANDARD; PRT; 985 AA.

AC P56526; O13451;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase).  
GN AGLA.

OS *Aspergillus niger*.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; *Aspergillus*.  
OX NCBI\_TaxID=5061;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GN-3;  
RX MEDLINE=97308536; PubMed=9165762;  
RA Nakamura A., Nishimura I., Yokoyama A., Lee D.-G., Hidaka M.,  
RA Masaki H., Kimura A., Chiba S., Uozumi T.;  
RT "Cloning and sequencing of an alpha-glucosidase gene from *Aspergillus*  
RT *niger* and its expression in *A. nidulans*.";  
RL J. Biotechnol. 53:75-84(1997).  
RN [2]  
RP SEQUENCE OF 26-252 AND 267-985.  
RC STRAIN=GN-8;  
RX MEDLINE=93005089; PubMed=1368849;  
RA Kimura A., Takata M., Sakai O., Matsui H., Takai N., Takayanagi T.,  
RA Nishimura I., Uozumi T., Chiba S.;  
RT "Complete amino acid sequence of crystalline alpha-glucosidase from  
RT *Aspergillus niger*.";  
RL Biosci. Biotechnol. Biochem. 56:1368-1370(1992).  
CC -!- FUNCTION: Hydrolyzes malto-oligosaccharides, but has a low  
CC activity toward soluble starch.  
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-  
CC linked D-glucose residues with release of D-glucose.  
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
DR EMBL; D45356; BAA3616.1; -;  
DR GlycosuiteDB; P56526; -;  
DR InterPro; IPR000322; Glyco\_hydro\_31.  
DR Pfam; PF01055; Glyco\_hydro\_31; 1.  
DR PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
DR PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.  
KW Hydrolase; Glycosidase; Glycoprotein; Signal.  
FT SIGNAL 1 25  
FT CHAIN 26 985 ALPHA-GLUCOSIDASE.  
FT ACT SITE 490 490  
FT CARBOHYD 36 36 O-LINKED (POTENTIAL).  
FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 143 143 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 218 218 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 347 347 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 422 422 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 506 506 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 537 537 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 545 545 O-LINKED (POTENTIAL).  
FT CARBOHYD 550 550 O-LINKED (POTENTIAL).  
FT CARBOHYD 559 559 O-LINKED (POTENTIAL).  
FT CARBOHYD 560 560 O-LINKED (POTENTIAL).  
FT CARBOHYD 561 561 O-LINKED (POTENTIAL).  
FT CARBOHYD 562 562 O-LINKED (POTENTIAL).  
FT CARBOHYD 601 601 N-LINKED (GLCNAC. . .).

FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .).	FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .).
FT CARBOHYD 835 835 N-LINKED (GLCNAC. . .).	FT CARBOHYD 835 835 N-LINKED (GLCNAC. . .).
FT CARBOHYD 881 881 N-LINKED (GLCNAC. . .).	FT CARBOHYD 881 881 N-LINKED (GLCNAC. . .).
FT CARBOHYD 899 899 N-LINKED (GLCNAC. . .).	FT CARBOHYD 899 899 N-LINKED (GLCNAC. . .).
FT CARBOHYD 957 957 N-LINKED (GLCNAC. . .).	FT CARBOHYD 957 957 N-LINKED (GLCNAC. . .).
FT CARBOHYD 970 970 N-LINKED (GLCNAC. . .).	FT CARBOHYD 970 970 N-LINKED (GLCNAC. . .).
FT VARIANT 27 28 TT -> LL (IN STRAIN GN-8).	FT VARIANT 27 28 TT -> LL (IN STRAIN GN-8).
FT VARIANT 42 42 D -> A (IN STRAIN GN-8).	FT VARIANT 42 42 D -> A (IN STRAIN GN-8).
FT VARIANT 929 929 N -> M (IN STRAIN GN-8).	FT VARIANT 929 929 N -> M (IN STRAIN GN-8).
SQ SEQUENCE 985 AA; 108913 MW; 9A18772AEB2E0927 CRC64;	SQ SEQUENCE 985 AA; 108913 MW; 9A18772AEB2E0927 CRC64;

Query Match 30.5%; Score 1410.5; DB 1; Length 985;  
Best Local Similarity 37.3%; Pred. No. 1.4e-93;  
Matches 306; Conservative 139; Mismatches 243; Indels 133; Gaps 22;

QY 59 DVQRLAVYASLETDSRLVRI-----TDADHPWB-VPODIIPRPAPGDVLHDAPPASAP 113	QY 59 DVQRLAVYASLETDSRLVRI-----TDADHPWB-VPODIIPRPAPGDVLHDAPPASAP 113
DB 95 DVESLTLSVEYQDSRLNIQLPHTVDSTNASWYFLSENLVPRP-----KAS 141	DB 95 DVESLTLSVEYQDSRLNIQLPHTVDSTNASWYFLSENLVPRP-----KAS 141
QY 114 LQGRVLSFAGSDLVLTVHASP-FRFTVSRSTGTLEDT-APGLVFRDKYLEVTSALPAG 171	QY 114 LQGRVLSFAGSDLVLTVHASP-FRFTVSRSTGTLEDT-APGLVFRDKYLEVTSALPAG 171
DB 142 LNASV---SQDOLFVSWNEPSFNFVKIRKATGALPSTEGTVLVYENQFIEFVTLPE- 197	DB 142 LNASV---SQDOLFVSWNEPSFNFVKIRKATGALPSTEGTVLVYENQFIEFVTLPE- 197
QY 172 RASLYGLGEHTKSSFLRHNDSTFLMNADIGASYVDVNLVYGHSHPFYMDVR----- 221	QY 172 RASLYGLGEHTKSSFLRHNDSTFLMNADIGASYVDVNLVYGHSHPFYMDVR----- 221
DB 198 EYNLYGLGEHI-TQFRLQRLNANLITPSDDGTP-IDQNLVGHQHPYLDTRYKGDQNGS 255	DB 198 EYNLYGLGEHI-TQFRLQRLNANLITPSDDGTP-IDQNLVGHQHPYLDTRYKGDQNGS 255
QY 222 -AP-----GTAHGVLLLSSNGMDVLGGSTVTKYVGGVLDFFPFGPNPLA 267	QY 222 -AP-----GTAHGVLLLSSNGMDVLGGSTVTKYVGGVLDFFPFGPNPLA 267
DB 256 YIPVKSSBADASQDYISLSHGVLNLSHGLEILLRSQKLIWRTLGGGIDLTLYSGPAPAD 315	DB 256 YIPVKSSBADASQDYISLSHGVLNLSHGLEILLRSQKLIWRTLGGGIDLTLYSGPAPAD 315
QY 268 VVDQY-TQLIARPAPMPYMSFGHCRCRYGLNVSDLERVARYAKARIPLEVMVTDIDYM 326	QY 268 VVDQY-TQLIARPAPMPYMSFGHCRCRYGLNVSDLERVARYAKARIPLEVMVTDIDYM 326
DB 316 VTRQLTSTVGLPAMQQTTLGFCRCRWGNNWSDADVAVANFEKFEIPLYMTDIDYM 375	DB 316 VTRQLTSTVGLPAMQQTTLGFCRCRWGNNWSDADVAVANFEKFEIPLYMTDIDYM 375
QY 327 DGFKDFTLDVNFYTAALRLPFVDRHRAQKYLILDPGIRV-DPIDAT--YGFVFRQM 383	QY 327 DGFKDFTLDVNFYTAALRLPFVDRHRAQKYLILDPGIRV-DPIDAT--YGFVFRQM 383
DB 376 HGRNFNDQHRFSYSEGDEFLSKLHESGRYVYPIVDAALYIPNPENASDAYAYDRGAA 435	DB 376 HGRNFNDQHRFSYSEGDEFLSKLHESGRYVYPIVDAALYIPNPENASDAYAYDRGAA 435
QY 384 QDIFLKR-NGTNFQNVGWPQGVYFPDMHAPAAEFWABEISLFRRTIPVDGLWIDMNBIS 442	QY 384 QDIFLKR-NGTNFQNVGWPQGVYFPDMHAPAAEFWABEISLFRRTIPVDGLWIDMNBIS 442
DB 436 DDVFLKNPDGSLYIGAVWPQYTVFDFWHKPAVDPMANELVMSKKVAPDGVVYDMSEVS 495	DB 436 DDVFLKNPDGSLYIGAVWPQYTVFDFWHKPAVDPMANELVMSKKVAPDGVVYDMSEVS 495
QY 443 NF-----YN-PEPMNA----- 452	QY 443 NF-----YN-PEPMNA----- 452
DB 496 SFCVSGCTGNLTNPAHPSPFLLPQEPGDIIYDPEAFNITNATEAASAGASSQAAAT 555	DB 496 SFCVSGCTGNLTNPAHPSPFLLPQEPGDIIYDPEAFNITNATEAASAGASSQAAAT 555
QY 453 -----LDDPPYRINNDGTGRPINNKTVRPLAVHYGGVTEYEHN 491	QY 453 -----LDDPPYRINNDGTGRPINNKTVRPLAVHYGGVTEYEHN 491
DB 556 ATTTSTSVSYLRTTTPGVRNVHEPPYVINHQBDLSVHASPNAHVDQGVVEYVHG 615	DB 556 ATTTSTSVSYLRTTTPGVRNVHEPPYVINHQBDLSVHASPNAHVDQGVVEYVHG 615
QY 492 LFGLEAPATGCVLR--DTGRRPFVLSRSTFVSGRYTAYTGDNAATWGLDLYSINTM 549	QY 492 LFGLEAPATGCVLR--DTGRRPFVLSRSTFVSGRYTAYTGDNAATWGLDLYSINTM 549
DB 616 LYGHQGLNATYQGLLEVMSHKRPFIIIGRSTFAGSGKWAGHMGDNYSKWMSYISQA 675	DB 616 LYGHQGLNATYQGLLEVMSHKRPFIIIGRSTFAGSGKWAGHMGDNYSKWMSYISQA 675
QY 550 LSPGLFGMPMIGADICGPNNGNTBELCGRMQLQCAFYFSDRHSATFTVRELYLWPSVA 609	QY 550 LSPGLFGMPMIGADICGPNNGNTBELCGRMQLQCAFYFSDRHSATFTVRELYLWPSVA 609
DB 676 LSPSLFGIPMFAGDTGCGFNGSDDELNRMWQLSAPFFTRHRELSTIPEPTRMASVI 735	DB 676 LSPSLFGIPMFAGDTGCGFNGSDDELNRMWQLSAPFFTRHRELSTIPEPTRMASVI 735
QY 610 ASGRKALGLRYQLLPYPVYTLMEYEAHTGAPIRLFFSYPHDVATYGVDRQLLGRGVLY 669	QY 610 ASGRKALGLRYQLLPYPVYTLMEYEAHTGAPIRLFFSYPHDVATYGVDRQLLGRGVLY 669
DB 736 EATKSAMRIRVAILPYPTLFDLAHTTGSTVMRALSWEFPNDPTLAAVETQFMVGPALMV 795	DB 736 EATKSAMRIRVAILPYPTLFDLAHTTGSTVMRALSWEFPNDPTLAAVETQFMVGPALMV 795
QY 670 SPVLEPGTTVDAYFPA---OR-WYRLYDYLAVATRTGKHVRLPAPADTVNVLGTGTI 725	QY 670 SPVLEPGTTVDAYFPA---OR-WYRLYDYLAVATRTGKHVRLPAPADTVNVLGTGTI 725
DB 796 VPVLEPLVNTYKGVFPGVGHEVYDWMYTOA-AVDAPGVNTTISAPLGHIPVTVVRGNI 854	DB 796 VPVLEPLVNTYKGVFPGVGHEVYDWMYTOA-AVDAPGVNTTISAPLGHIPVTVVRGNI 854
QY 726 LPLQSSALTTSSRRTAFHLLVALAEDGTASGYLPDDGDS 766	QY 726 LPLQSSALTTSSRRTAFHLLVALAEDGTASGYLPDDGDS 766
DB 855 LPMQEPALTTREARQTPWALLAALGNSGTASGQYLDDGES 895	DB 855 LPMQEPALTTREARQTPWALLAALGNSGTASGQYLDDGES 895

RESULT 6

LYAG HUMAN  
ID \_LYAG HUMAN STANDARD; PRT; 952 AA.  
AC P10253; Q14351; Q16302;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Lysosomal alpha-glucosidase precursor (EC 3.2.1.20) (Acid maltase).  
GN GAA.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE OF 70-89; 123-145; 204-215;  
RP 230-249; 332-345; 349-370; 394-409; 480-513; 520-545; 703-719; 726-731  
RP AND 799-803.  
RC TISSUE=Placenta, Testis, and Urine;  
RX MEDLINE=89005058; PubMed=3049072;  
RA Hoefsloot L.H., Hooogeveen-Westerveld M., Kroos M.A., van Beeumen J.,  
RA Reuser A.J.J., Oostra B.A.;  
RT "Primary structure and processing of lysosomal alpha-glucosidase;  
RT homology with the intestinal sucrase-isomaltase complex.";  
RL EMBO J. 7:1697-1704(1988).  
[2]  
RP REVISIONS.  
RA Reuser A.J.J.;  
RL Submitted (JUN-1990) to the EMBL/GenBank/DBJ databases.  
[3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90262651; PubMed=2111708;  
RA Martiniuk F., Mehler M., Tzall S., Meredith G., Hirschhorn R.;  
RT "Sequence of the cDNA and 5'-flanking region for human acid alpha-  
RT glucosidase. detection of an intron in the 5' untranslated leader  
RT sequence, definition of 18-bp polymorphisms, and differences with  
RT previous cDNA and amino acid sequences.";  
RL DNA Cell Biol. 9:85-94(1990).  
[4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91097465; PubMed=2268276;  
RA Hoefsloot L.H., Hooogeveen-Westerveld M., Reuser A.J.J., Oostra B.A.;  
RT "Characterization of the human lysosomal alpha-glucosidase gene.";  
RL Biochem. J. 272:493-497(1990).  
[5]  
RP ACTIVE SITE.  
RX MEDLINE=91310614; PubMed=1856189;  
RA Hermans M.M.P., Kroos M.A., van Beeumen J., Oostra B.A.,  
RA Reuser A.J.J.;  
RT "Human lysosomal alpha-glucosidase. Characterization of the catalytic  
RT site.";  
RL J. Biol. Chem. 266:13507-13512(1991).  
[6]  
RP CARBOHYDRATE-LINKAGE SITES.  
RX MEDLINE=93168114; PubMed=8435067;  
RA Hermans M.M.P., Wisselaar H.A., Kroos M.A., Oostra B.A.,  
RA Reuser A.J.J.;  
RT "Human lysosomal alpha-glucosidase: functional characterization of  
RT the glycosylation sites.";  
RL Biochem. J. 289:681-686(1993).  
[7]  
RP REVIEW ON VARIANTS.  
RX MEDLINE=95327152; PubMed=7603530;  
RA Reuser A.J.J., Kroos M.A., Hermans M.M.P., Bijvoet A.G.A.,  
RA Verbeet M.P., van Diggelen O.P., Kleijer W.J., van der Ploeg A.T.;  
RT "Glycogenosis type II (acid maltase deficiency).";  
RL Muscle Nerve 3:S61-S69(1995).  
[8]  
RP VARIANT ASN-91.  
RX MEDLINE=90365036; PubMed=2203258;  
RA Martiniuk F., Bodkin M., Trall S., Hirschhorn R.;  
RT "Identification of the base-pair substitution responsible for a human  
RT acid alpha glucosidase allele with lower 'affinity' for glycogen (GAA  
RT 2) and transient gene expression in deficient cells.";  
RL Am. J. Hum. Genet. 47:440-445(1990).

[9]  
RN VARIANT GSD-II THR-318.  
RP MEDLINE=91353580; PubMed=1652892;  
RX Zhong N., Martiniuk F., Tzall S., Hirschhorn R.;  
RT "Identification of a missense mutation in one allele of a patient  
RT with Pompe disease, and use of endonuclease digestion of  
RT PCR-amplified RNA to demonstrate lack of mRNA expression from the  
RT second allele.";  
RL Am. J. Hum. Genet. 49:635-645(1991).  
[10]  
RN VARIANT GSD-II LYS-521.  
RP MEDLINE=91379015; PubMed=1898413;  
RX Hermans M.M.P., de Graaff E., Kroos M.A., Wisselaar H.A., Oostra B.A.,  
RA Reuser A.J.J.;  
RT "Identification of a point mutation in the human lysosomal alpha-  
RT glucosidase gene causing infantile glycogenosis type II.";  
RL Biochem. Biophys. Res. Commun. 179:919-926(1991).  
[11]  
RN VARIANTS GSD-II ARG-643 AND TRP-725.  
RX MEDLINE=94004908; PubMed=8401535;  
RA Hermans M.M.P., Kroos M.A., de Graaff E., Oostra B.A., Reuser A.J.J.;  
RT "Two mutations affecting the transport and maturation of lysosomal  
RT alpha-glucosidase in an adult case of glycogen storage disease type  
RT II.";  
RL Hum. Mutat. 2:268-273(1993).  
[12]  
RN VARIANT GSD-II GLU-645, AND VARIANTS ILE-816 AND ILE-927.  
RX MEDLINE=93168115; PubMed=8094613;  
RA Hermans M.M.P., de Graaff E., Kroos M.A., Wisselaar H.A.,  
RA Willemsen R., Oostra B.A., Reuser A.J.J.;  
RT "The conservative substitution Asp-645-->Glu in lysosomal alpha-  
RT glucosidase affects transport and phosphorylation of the enzyme in an  
RT adult patient with glycogen-storage disease type II.";  
RL Biochem. J. 289:687-693(1993).  
[13]  
RN VARIANT GSD-II GLU-645, AND VARIANTS ILE-816 AND ILE-927.  
RX MEDLINE=92096118; PubMed=1684505;  
RA Martiniuk F., Mehler M., Bodkin M., Tzall S., Hirschhorn K., Zhong N.,  
RA Hirschhorn R.;  
RT "Identification of a missense mutation in an adult-onset patient with  
RT glycogenosis type II expressing only one allele.";  
RL DNA Cell Biol. 10:681-687(1991).  
[14]  
RN VARIANTS ILE-816 AND ILE-927.  
RX MEDLINE=93252406; PubMed=8486380;  
RA Hermans M.M.P., Svetkey L.P., Oostra B.A., Chen Y.T., Reuser A.J.J.;  
RT "The loss of a polymorphic glycosylation site caused by Thr-927-->Ile  
RT is linked to a second polymorphic Val-816-->Ile substitution in  
RT lysosomal alpha-glucosidase of American blacks.";  
RL Genomics 16:300-301(1993).  
[15]  
RN VARIANT GSD-II VAL-519.  
RX MEDLINE=95170739; PubMed=7866409;  
RA Huie M.L., Hirschhorn R., Chen A.S., Martiniuk F., Zhong N.;  
RT "Mutation at the catalytic site (M519V) in glycogen storage disease  
RT type II (Pompe disease).";  
RL Hum. Mutat. 4:291-293(1994).  
[16]  
RN VARIANT GSD-II TRP-647.  
RX MEDLINE=95072571; PubMed=7981676;  
RA Huie M.L., Chen A.S., Brooks S.S., Grix A., Hirschhorn R.;  
RT "A de novo 13 nt deletion, a newly identified C647W missense mutation  
RT and a deletion of exon 18 in infantile onset glycogen storage disease  
RT type II (GSDII).";  
RL Hum. Mol. Genet. 3:1081-1087(1994).  
[17]  
RN VARIANT GSD-II LEU-545.  
RX MEDLINE=95187163; PubMed=7881422;  
RA Hermans M.M.P., de Graaff E., Kroos M.A., Mohkamsing S., Eussen B.J.,  
RA Joosse M., Willemsen R., Kleijer W.J., Oostra B.A., Reuser A.J.J.;  
RT "The effect of a single base pair deletion (delta 1525) and a C1634T  
RT missense mutation (Pro545Leu) on the expression of lysosomal alpha-  
RT glucosidase in patients with glycogen storage disease type II.";

RL Hum. Mol. Genet. 3:2213-2218 (1994).  
 RN [18]  
 RP VARIANTS GSD-II ARG-299; LYS-903 DEL AND VARIANTS HIS-199; ARG-223 AND  
 RX VAL-780.  
 RA MEDLINE=95233437; PubMed=7717400;  
 RA Boerkoel C.F., Exelbert R., Nicastri C., Nichols R.C., Miller F.W.,  
 RA Plotz P.H., Raben N.;  
 RT "Leaky splicing mutation in the acid maltase gene is associated with  
 RT delayed onset of glycogenosis type II.";  
 RL Am. J. Hum. Genet. 56:887-897 (1995).  
 RN [19]  
 RP SEQUENCE OF 631-680 FROM N.A., AND VARIANT GSD-II HIS-645.  
 RX MEDLINE=95209708; PubMed=7695647;  
 RA Lin C.-Y., Shieh J.-J.;  
 RT "Identification of a de novo point mutation resulting in infantile  
 RT form of Pompe's disease.";  
 RL Biochem. Biophys. Res. Commun. 208:886-893 (1995).  
 RN [20]  
 RP VARIANT GSD-II VAL-529.  
 RX MEDLINE=96431168; PubMed=8834250;  
 RA Tsunoda H., Ohshima T., Tohyama J., Sasaki M., Sakuragawa N.,  
 RA Martinik F.;  
 RT "Acid alpha-glucosidase deficiency: identification and expression of  
 RT a missense mutation (S529V) in a Japanese adult phenotype.";  
 RL Hum. Genet. 97:496-499 (1996).  
 RN [21]  
 RP VARIANTS GSD-II ASN-645; TRP-647; SER-648; GLN-672 AND TRP-672.  
 RX MEDLINE=98205825; PubMed=9535769;  
 RA Huie M.L., Tajino S., Brooks S.S., Engel A., Elias E., Bonthron D.T.,  
 RA Bessler C., Shanks S., Damauro S., Goto Y.I., Hirschhorn R.;  
 RT "Glycogen storage disease II: identification of four novel  
 RT missense mutations (D645N, G648S, R672W, R672Q) and two  
 RT insertions/deletions in the acid alpha-glucosidase locus of patients  
 RT of differing phenotype.";  
 RL Biochem. Biophys. Res. Commun. 244:921-927 (1998).  
 RN [22]  
 RP VARIANTS GSD-II PRO-566; ARG-643 AND ARG-768, AND VARIANTS ASN-91;  
 RX HIS-199 AND ARG-223.  
 RA MEDLINE=98180719; PubMed=9521422;  
 RA Hermans M.P.P., Kroos M.A., Smeitink J.A.M., van der Ploeg A.T.,  
 RA Kleijer W.J., Reuser A.J.J.;  
 RT "Glycogen storage disease type II: genetic and biochemical analysis  
 RT of novel mutations in infantile patients from Turkish ancestry.";  
 RL Hum. Mutat. 11:209-215 (1998).  
 RN [23]  
 RP VARIANT GSD-II GLY-VAL-PRO-SER-ASN-925 INS.  
 RA Beesley C.E., Child A.H., Yacoub M.Y.;  
 RT "The identification of five novel mutations in the lysosomal acid  
 RT alpha-(1,4) glucosidase gene from patients with glycogen storage  
 RT disease type II.";  
 RL Hum. Mutat. 11:413-413 (1998).  
 RN [24]  
 RP VARIANT GSD-II ARG-481.  
 RX MEDLINE=99202470; PubMed=10189220;  
 RA Raben N., Lee E., Lee L., Hirschhorn R., Plotz P.H.;  
 RT Query Match 30.3%; Score 1401.5; DB 1; Length 952;  
 RT Best Local Similarity 37.2%; Pred. No. 6.1e-93; Gaps 25;  
 RT Matches 330; Conservative 134; Mismatches 327; Indels 95; Gaps 25;  
 QY 14 CLFAPRLCSKEGFLAARTVLAVALVMEGALRAEAATGCRSTGQVQRLAVVASLETDS 73  
 DB 127 CFFPPSPVSKLENLSSEMCYTALT-----RTTFFPKDILTLDVMMETEN 177  
 QY 74 RLVRITADHPREVEQDIIIPRAPGDLVHDAPPASSAPLOGRVLSFAGSDLVLTWTHAS 133  
 DB 178 RLHFTIKDPANRRYEVPLE-TPR-----VHSRAP-----SPLYSVFFS-----EE 216  
 QY 134 PPRFTVSRRTGDTLFTD--APGLVFRDKYLEVTSALPAGRASLYGIGETHKTSFRLRH 191  
 DB 217 PFGVIVIRQLDGRVLLNTVAP-LFFADQFLQLSTLPS--QYITGLAEHLSPMLNSTSW 273  
 QY 192 DSFTLWADIGASYVDVNLVYGHSPFYMDVRAPGTAHGVLLSSNGMD-VLYGGSYVTYKV 250

Db 274 TRITLWNRDL-APTGANLYGSHFFYLALEGGSAHGVFLNSNMDVVLOPSPALSWSRS 332  
 QY 251 IGGVLDFYFFAGNPPLAVVQYQTLIARPAFMPYMSFGFHQCRYGYLNVSDLRVARVYA 310  
 Db 333 TGGILDVYIFLGPSPKSVVQYLDVGVYFPPFPPYGLGFLHCRWGSYSTAITROVVENMT 392  
 QY 311 KARPLEVMTDIDYMGFKDFTLDRVNPTAAELRPPFVDRLRHNAQKVLLIDRCIRVDP 370  
 Db 393 RAHFPLDVQMDLDYMSRRDFTFNKOGF--RDPFAMVQELHQGRRYMMIVLDPFAISSG 450  
 QY 371 IDATYGTFRGMQODIFL-KKNGTNFVGNVWPGDVYFPDFMHPAAAEFWAEISLFRRTI 429  
 Db 451 PAGSVRPYDEGLRGVFTNETGOLICKVWPGSTAFDFNTPTALAWEDMVAEFHQDV 510  
 QY 430 PVDGLWIDMNEISFYNPE-----PMNALDDPPYRINNDGTGRPINNKTVRFLAVHYGVT 485  
 Db 511 PFDGMWIDMNEPSNPIRGSEDCPCNNELENPPY--VPGVVGSTLQAATI-CASSHQFLST 567  
 QY 486 EYEHNLFGLEAPATGRGLRDTGRPPFLVSRSTFVSGSRGTAYTGTGDNATWCDLRS 545  
 Db 568 HYNLNLVGLTEATASHRALVKARGTRPPVISRSSTFAGHRYAGHMTGDVWSSWEQLASS 627  
 QY 546 INTMLSFGLFGPMIGADICGFNGNNTTEELCGRWITLQAGFYPPSRDHSIAITVRRELYW 605  
 Db 628 VPEILQFNLLGVPLVGVADVCGFLGNTSEELCVRWITLQAGFYPPFMENHNSLLSLPOEYSF 687  
 QY 606 PSVAASG-RKALGLRYQLLPYFTLYMAEHTAGIAPRLPFSYPHDVATYGVDRQFLG 664  
 Db 688 SEPAQANRKALTLYALLPHLYTLFHOAHVAGETVARPLFLFFPKDSSTWTVHQHLLWG 747  
 QY 665 RGVLSVPLEPGPTVDAFYPAGRWYRLDYSL-----AVATR-----TGKHVRL 709  
 Db 748 EALLTPVQAKAEVTCYFPLGTWYDLOTVPIEALGSLPPPPAPRPAHSEGOVWTL 807  
 QY 710 PAPADTVNVHLTGGTTLPLQOSALTTTSRARTAFHLLVALAEDGTASGYLFDGDSPEY 769  
 Db 808 PAPLDTINVHLRAGYIIPLOGPLTTTTSRQOPMALAVALTKGGEARGLFWDDGESLEV 867  
 QY 770 GRSQWSVRF---NYKIPNNKGAIKVSEVHNSYAOSRTLVISKVVLVGHRSPPAPPK 826  
 Db 868 LERGAYTOVIFLARNTTVNE--LVRTSE-----GAGLQLOKQVTLG--VATAPQQ 915  
 QY 827 LTVHNSAEVEASS---SAGTRYONAGGLGGVAHIGLSLVVGEF 869  
 Db 916 ----VLSNGVPVSNFTYSPDKYLDI-----CVSLLMGEQF 947  
 RESULT 7  
 LYAG MOUSE  
 ID LYAG MOUSE STANDARD; PRT; 953 AA.  
 AC P70659;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Lysosomal alpha-glucosidase precursor (EC 3.2.1.20) (Acid maltaase).  
 GN GAA.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/c; TISSUE=Liver;  
 RL Ding J.H., Yang B.Z., Reuser A.J.J., Roe C.R.;  
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: ESSENTIAL FOR THE DEGRADATION OF GLYCOGEN TO GLUCOSE IN  
 CC LYCOSOMES.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-  
 CC linked D-glucose residues with release of D-glucose.  
 CC -!- SUBCELLULAR LOCATION: Lysosomal.  
 CC -!- SIMILARITY: Contains 1 P-type (trefoil) domain.  
 CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [licences@ebi.ac.uk](mailto:licences@ebi.ac.uk)).

DR EMBL: U49351; AB06943.1; -;  
 DR HSP: P01359; 2PSP  
 DR MGD; MGI:95609; Gaa.  
 DR InterPro; IPR000322; Glyco\_hydro\_31.  
 DR InterPro; IPR000519; P\_trefoil.  
 DR Pfam; PF01055; Glyco\_hydro\_31; 1.  
 DR Pfam; PF00088; trefoil; 1.  
 DR SMART; SM00018; PD; 1.  
 DR PROSITE; PS00025; P\_TREFOIL; 1  
 DR PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
 DR PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.  
 DR Hydrolase; Glycosidase; Glycoprotein; Lysosome; Signal.  
 FT SIGNAL 1 27 POTENTIAL.  
 FT PROPEP 28 69 BY SIMILARITY.  
 FT CHAIN 70 953 LYSOSOMAL ALPHA-GLUCOSIDASE.  
 FT DOMAIN 81 130 P-TYPE.  
 FT ACT\_SITE 518 518 BY SIMILARITY.  
 FT DISULFID 82 109 BY SIMILARITY.  
 FT DISULFID 92 108 BY SIMILARITY.  
 FT DISULFID 103 127 BY SIMILARITY.  
 FT CARBOHYD 140 140 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 390 390 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 470 470 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 883 883 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 926 926 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 933 933 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 953 AA; 106168 MW; 927EE0C7E1D59571 CRC64;

Query Match 30.3%; Score 1401; DB 1; Length 953;  
 Best Local Similarity 38.1%; Pred. No. 6.7e-93;  
 Matches 327; Conservative 122; Mismatches 323; Indels 86; Gaps 20;

QY 14 CLFAPRLCSKEEGLAARTVLAVATMEGALRAEAATCGRSTGQVORLAVYASLETTDS 73  
 DB 127 CFPPSPYSVRLN-----LSSTESSYATLTRTSTPTFPKDLTLQLEVLMEIDS 177  
 QY 74 RLVRITDADHPFWEVPQDIIRPAPGVLDHPAPASSAPLOGRVLSFAGSDLVLTVAAS 133  
 DB 178 RLHFRIKDPASKRYEYPLE-TPR-----VLSQAP-----SPLYSVEFS-----EE 216  
 QY 134 PRFTVSRSTGDTLFDLT--APGLVPRDKYLEVTSALPAGRASLVGLGHEHTKSSFLRHN 191  
 DB 217 PFGVIVRKLGGVRLNLTIVAP-LFADQFLQLSTSLPA--QHITGLGHEHLSPLMLSTDW 273  
 QY 192 DSTLNLADIGASYVDNLYGSHFFVMDVRAGTAHGVLLSSNGMDVLYGGS-VYTVKV 250  
 DB 274 ARTLNLNRDTPSQ-CTNLYGSHFFVLALEDGGLAHGVFLNLSNAMDVILQSPALTWRS 332  
 QY 251 IGVGLDFYFAGNPLAVVDQYQLIARPAHPYWSFGHCQRYGVLYNVSLERVARYA 310  
 DB 333 TGGILOVYFLGPEPKSVQOYLDVGVYPPMPYMGVGLFHLRCRWSSYTAIRVQVEMNT 392  
 QY 311 KARIPLEVMWTDIDYMDGPFDTLDRVNETAAELRPFVDRLHRAKQYVLLDPGIRVDP 370  
 DB 393 RTHFFPLDVQWLDYMDARDFTFNQDSF--ADFPDMVRDVHVGGRYVMWIDPAISSAG 450  
 QY 371 IDATYFTFVGMQDDIPL-KRGNTFVGNVWPGDVTFFDFMIPAAAFWARISLFRRTI 429  
 DB 451 PAGSYRPEDECLRGVFTNETGOPLIGKYCPGTTAFPDFTNPETLDMWQDMVSEFHXQV 510  
 QY 430 PVGLMTDMNEISNFTNPE----PMALDDPPYRINNDGTGRPINNKTVPLAVHYGGVT 485  
 DB 511 PFDGMWLDWNEPSNFRVGSQGGCPNNELENPY---VPGVUGILQAAICASSHQFLST 567

## RESULT 8

AGLU SCHPO STANDARD; PRT; 969 AA.  
 ID AGLU SCHPO  
 AC Q9C014;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase).  
 GN AGLU OR SPAPB24D3.10C  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 CX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A. SEQUENCE OF 25-36; 175-194; 375-395 AND  
 RP 427-451, AND MUTAGENESIS OF ASP-481; GLU-484 AND ASP-647.  
 RX MEDLINE=21195240; PubMed=11298744;  
 RA Okuyama M., Okuno A., Shimizu N., Mori H., Kimura A., Chiba S.;  
 RT "Carboxyl group of residue Asp647 as possible proton donor in  
 RT catalytic reaction of alpha-glucosidase from Schizosaccharomycetes  
 RT pombe.";  
 RL Eur. J. Biochem. 268:2270-2280(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sproules J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Muncall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,





[illegible]

FT	ACT SITE	462	462	BY SIMILARITY.
FT	DOMAIN	519	532	SER/THR-RICH.
FT	CARBOHYD	51	51	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	68	68	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	97	97	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	187	187	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	244	244	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	373	373	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	393	393	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	406	406	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	437	437	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	505	505	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	570	570	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	772	772	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	801	801	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	895	895	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	912	912	N-LINKED (GLCNAC. .) (POTENTIAL).
SQ	SEQUENCE	946 AA;	105804 MW;	BD6B640C4EEF1F70 CRC64;
Query March 29.5t; Score 1366.5; DB 1; Length 946;				
Best Local Similarity 33.2t; Pred. No. 2e-90;				
Matches 327; Conservative 145; Mismatches 323; Indels 191; Gaps 28;				
Qy	17	APRLCSSKESGKGLAARTVLAVATM-----EGALRAEAATGG-----RSSGT--	58	
Db	21	APTSSSAEEAQTVPVLSIGVKQLPNIHNSAVDNAVAKGSLVNVSLTARGLTGIL	80	
Qy	59	-----DVORLAVASLETDSRLVRITDAD-HPRWEVPQDIIPRP-APGDVLHD	105	
Db	81	KLKEATNIYGDFEYLNLSVEYOSDTPLNHIETDLTVFLPEELSVVAPKLEGD----	136	
Qy	106	APPASSAPLQGRVLSPAGSLDLTVHASPRFTVSRBSTGDTLFDTPAG-LVPRDKYLEV	164	
Db	137	-----AKTFNFENSLDVEYDEEDGFVLSRSSTREVLFTKGNPLVFSNQFIQF	186	
Qy	165	TSALPAGRASLYGLGEHTKTSFRLRHNS---FTLMNADIGASYVDVNLXGSHPFYMDVR	221	
Db	187	NTTLPKGH-SITGLGESIHGSL-----NEPGWKTLYANDI-ADPIDGNIYGVHPVYDQR	240	
Qy	222	APCTAHGVLLLSNGMDVLVYGGSYTYKVGVLDFYFFAGNPPLAVDQYQTOLIARPA	280	
Db	241	YDTNTHGVYTWRTSAIQEVVVGSTSLTWRALSGLVDILYFFSGPDPKDVIQOYVSEIGLPA	300	
Qy	281	PMPYWSGFFHQCRYGYLNVSDLERVVARAKAPILEVMMTDDYMDGFKDFTLDRVNTF	340	
Db	301	MQPYWALGYHOCRWGYDTVESLETVTVENFKKDIPLETIWSDDIDYMDGYKDFNDPVPFP	360	
Qy	341	AAELRPFVDRLHRNAQXVILDPGRIV-DPIDAT---YCTFVRGMQODFLER-NGTNF	395	
Db	361	TDKFKFLDLHNNSQHYVPFIQAAIYVYPNPATDDDYEPFLHGNESDFVLKPNQDGLS	420	
Qy	396	VGNVPQGVYFPQFMHPAAAEFWAREISLSPRTIPVDGLMIDMNEISNFY-----	445	
Db	421	IGAWPQYTVPPDFLANNTQGYWNKFKQWYERIPFDGIWTDNWEVSSFCVSGCGTDYRF	480	
Qy	446	-NP-----EPWNA-----	452	
Db	481	DNPVHPFPFVGYSGSDYPLGFDKSNASEMKSISEAAAAATKTTTSSASTSIDGKNTLA	540	
Qy	453	-----LDDPYRINNDGTGRPINNKTVPRLAVHYGVTEYEENLFGLLPARATGRGVLR	507	
Db	541	PGKGNINYPPIANNOGDHDLATHAISPNATHADGTVEYDIHNIYGLIOERAIYRALLE	600	
Qy	508	-DTGRPFVLSRSTFVSGRYTAIWTGDNAATWGLDIRYSINTMLSFGLQFMPIGADICG	566	
Db	601	IHPMKRPFIIIGRSSFAGSGKYMHWGNDYADYVMYVESIPQALSMGLSGIPFGVDACG	660	
Qy	567	PNGNTTEELCGRMVQLGAFYPPSRDHSIAFTVRELYPLWPSVAASGRKALGLRYQLLPYF	626	
Db	661	FNGNTDMELCSRWQLASFFPPFYRNHNVLGAIPQEPYVWEAVMKATKTSINVRYSLLPY	720	
Qy	627	YTLMYEAHMTGAPIARPLFFSYPHDVATYGVDRQFLLGRGLVSPVLEPGPITTVDAVFP	686	

Db 721 YTLHSHVGTGIPIMRAFNWQPPYSKELAGVDTOFVGDALLVTPVLEPGVNNHTKGFPG 780  
 Qy 687 GR - WRLVYDYLAVATRGKHVRLPAPADTVNVHLTGTGTLPLQOQSALTTSRARTAFH 744  
 Db 781 ENAVYDFTHKKQFT - AGNETLAAPLCHPLIKGNLIPTQEPGYTTTSSKRPFG 839  
 Qy 745 LLVALAABDGTASGYLFLDDGDSPEYGRSDMSVRENYKIPNNKGAIKVSEVVHNSYQA 804  
 Db 840 LLVALDAEGTAGSKLYLDGGSVDV - - - - - EALVVDVFA 874  
 Qy 805 STLVIKVVLMCH - - - - - RSPAA - - - - - PKLTVHNSAEVEASSSAGTYQVAGG 851  
 Db 875 SKNKLAVSV - FGEYVROPLANVTILGVDSFPPKVLFN - - - - - NETVSHKYEN - - - 921  
 Qy 852 LGCVAHIGGLSL - - - - - VVGEFFELK 872  
 Db 922 - GAVYLTDLKFTKEGAPAEFTLQ 945

RESULT 11  
 AGLU ASPOR  
 ID AGLU ASPOR STANDARD; PRT; 985 AA.  
 AC Q12558;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase) (AGL).  
 GN AGDA.  
 OS Aspergillus oryzae.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 OX NCBI\_Taxid=5062;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RIB 40;  
 RX MEDLINE=96032211; PubMed=7549103;  
 RA Minetoki T., Gomi K., Kitamoto K., Kumagai C., Tamura G.;  
 RT "Nucleotide sequence and expression of alpha-glucosidase-encoding  
 gene (agda) from Aspergillus oryzae."  
 RL Biosci. Biotechnol. Biochem. 59:1516-1521(1995).  
 CC -!- FUNCTION: Hydrolyzes malto-oligosaccharides, but has a low  
 CC activity toward soluble starch.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-  
 CC linked D-glucose residues with release of D-glucose.  
 CC -!- INDUCTION: By maltose.  
 CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; D45179; BAA08125.1; -  
 CC PIR; JC4217; JC4217.  
 CC InterPro: IPR000322; Glyco\_hydro\_31.  
 CC Pfam; PF01055; Glyco\_hydro\_31; 1.  
 CC PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
 CC PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.  
 CC Hydrolyase; Glycosidase; Glycoprotein; Signal.  
 CC SIGNAL 1 25  
 CC BY SIMILARITY.  
 CC CHAIN 26 985  
 CC ALPHA-GLUCOSIDASE.  
 CC ACT SITE 492 492  
 CC BY SIMILARITY.  
 CC FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 145 145 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 220 220 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 255 255 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 349 349 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 508 508 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 536 536 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 539 539 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 602 602 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 624 624 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 661 661 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 835 835 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 881 881 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 929 929 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 957 957 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 985 AA; 108704 MW; 3E9EAB0AE38209E CRC64;  
 Query Match 29.1%; Score 1348.5; DB 1; Length 985;  
 Best Local Similarity 34.1%; Pred. No. 4.2e-89;  
 Matches 312; Conservative 158; Mismatches 289; Indels 157; Gaps 28;  
 Qy 59 DVORLAVASLETDSRLVRIT - - - - - DADHPRWEV - PQDIIIPRAPGDVLDHDPASSAP 113  
 Db 97 DVDSLTLTVYQAKDRLNIQIVPTTFDASNASWYILSEELVRP - - - - - KASQNASVP 149  
 Qy 114 LOGRVLSPAGSDLVLTVHASP - FRFTVSRRTSGDTLFD - APGLVFRKYLEVTSALPAG 171  
 Db 150 - - - - - QSDFFVVSWSNPFNFVKIRKATGDVLENTKSTLVYENQFIEFVTLPE - 199  
 Qy 172 RASLYGLGHTKSSFLRHNDSEFTLNADIGASYVDVNLVYGHSPYMDVR - - - - - 221  
 Db 200 EYNLYGLGER - MNQLRLLENAMLTLYAADI - ADPIDDNIYGHAFYLTTRYKYVGQONKS 257  
 Qy 222 - - - - - APGT - - - - - AHGVLLSSNGMDVLYGGSVTVKVIYGVLDFFVFFAGNPILA 267  
 Db 258 HTIVKSSEAEPSQEVVSYSHGVFLRNAHQEILLRDKLIWRTLGGSVDLTFYSGPTQAE 317  
 Qy 268 VVDQV - TQLIAPAPMPYMSFGHCYGYLNVSDLERVARYAKARIPLEVMVTDIDYM 326  
 Db 318 VTKQYOLSTVGLPAMQYNTLGFHQCRWYNNWSEFEDVLANFERFEIPLVLEWADIYM 377  
 Qy 327 DGFKQFTLDRVNTAAELRPFVDRLRHNAQKYLVLDPGIRV - DPIDAT - - YGTFVRGMQ 383  
 Db 378 HGVRPNFDNDQHRFSYEEGEKFLKLHAGGRVPIVDGALYIPNENASDAYETVDRGAK 437  
 Qy 384 QDIFLKR - NGTMFVGNVPGDYVDFDMPHAPAAEFWABEISLFRRTIPVDGLWIDMNEIS 442  
 Db 438 DDVFIKNPDGSLYIGAVMPGYTVYPDWHHPKASDPWANELVTMMNKLHYDGVVYDMAEVS 497  
 Qy 443 NF - - - - - YNP - - - - - EPNNA - - - - - 452  
 Db 498 SFCVSGCTGNLSMNPAPHPALPGEPGVNVVDYEGFNITNATEASASASQAASAA 557  
 Qy 453 - - - - - LDDPPYRINNDGTGRPINNKTVRPLAVHYGGVTEENHL 492  
 Db 558 SSTTSAPYLRTPTPGVRNVDPHPVYVNHVQGHDLVSHAISPSNSTHSDGVQYDVHSL 617  
 Qy 493 FGLLEARATGRGLRD - TGRPRPVLRSFTVSGRYTAYWTCGDNAATWGLDLYSINTMLS 551  
 Db 618 YHQGINATVHGLLKVWENKRFPIARSTFGSGKMGHWGDNFSKMGSMFFSISQALQ 677  
 Qy 552 FGLFGPMIADICGPNNGNTTEELCGRWIQLGAPYFPFSRDHSAIFTVRELYLWPSVAAS 611  
 Db 678 FSLFGIPMGVDTGCGNGNTDEELCNRMNQLSAFPFPFYRNHNVLSAIPQEPYRNASVIDA 737  
 Qy 612 GRVALGRYQLLPYFYTLMYEAMTICAPARPLFFSYPHDVATYGVDRQFLGLGRVLSVP 671  
 Db 738 TKAAAMIRYAILPFYTLFHLAHTFTVMRALAWEFNDPSLAAGVGTQFLVGPSPVMVIP 797  
 Qy 672 VLEPGPTTVDAVFP - - - - - GR - WYRLYDYSLAVATRTGKHVRLPAPADTVNVHLTGTIIP 727  
 Db 798 VLEPQVDTVQGVPGVGHGEVNDWYSQT - AVDAKEGVNNTTISAPLGHIPVFRVGSILP 856  
 Qy 728 LQOQSALTTSRARTAFHLLVALAEDGTASGYLFLDDGDS - - PEYGRSDMSVMRFYKIP 785  
 Db 857 MQEVALTTDARKTPTWSILLASLSSNGTASQYLDGGSVYPEDTLSDVFLASRSTLR - A 915  
 Qy 786 NNKGALVKSEVVHNSYASRTLVISKVVLMGHRSAPAPKLTIVHNSNEVEASSAGTR 845  
 Db 916 SARGTWKEANP - - - - - LANVTVLG - - - - - VTEKPSVSVTLNG - - - - - ETLSDSVYK 955

QY 846 YQNAGGLGGVAHIGGL 861  
 DB 956 Y---NATSHVLVGL 968

RESULT 12  
 SUIIS HUMAN  
 ID SUIIS HUMAN STANDARD; PRT; 1826 AA.  
 AC P14410;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Sucrase-isomaltase, intestinal (Contains: Sucrase (EC 3.2.1.48);  
 DE Isomaltase (EC 3.2.1.10)).  
 GN SI.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Intestine;  
 RX MEDLINE=92359963; PubMed=1353958;  
 RA Chantret I., Lacasa M., Chevalier G., Ruf J., Islam I., Mantel N.,  
 RA Edwards Y., Swallow D., Rousset M.;  
 RT "Sequence of the complete cDNA and the 5' structure of the human  
 RT sucrase-isomaltase gene. Possible homology with a yeast  
 RT glucosylase";  
 RL Biochem. J. 285:915-923 (1992).  
 RN [2]  
 RP SEQUENCE OF 1-677 FROM N.A.  
 RX MEDLINE=88112852; PubMed=2962903;  
 RA Green F., Edwards Y., Hauri H.-P., Povey S., Ho M.W., Pinto M.,  
 RA Swallow D.;  
 RT "Isolation of a cDNA probe for a human jejunal brush-border  
 RT hydrolase, sucrase-isomaltase, and assignment of the gene locus to  
 RT chromosome 3";  
 RL Gene 57:101-110 (1987).  
 RN [3]  
 RP VARIANT DISACCHARIDE INTOLERANCE I PRO-1097.  
 RX MEDLINE=96189940; PubMed=8609217;  
 RA Ouwendijk J., Moolenaar C.E.C., Peters W.J., Hollenberg C.P.,  
 RA Ginsel L.A., Franssen J.A.M., Naim H.Y.;  
 RT "Congenital sucrase-isomaltase deficiency: identification of a  
 RT glutamine to proline substitution that leads to a transport block of  
 RT sucrase-isomaltase in a pre-Golgi compartment";  
 RL J. Clin. Invest. 97:633-641 (1996).  
 CC -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE FINAL STAGE OF  
 CC CARBOHYDRATE DIGESTION.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of sucrose and maltose by an  
 CC alpha-D-glucosidase-type action.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,6-alpha-D-glucosidic linkages  
 CC in some oligosaccharides produced from starch and glycogen by  
 CC alpha-amylase, and in isomaltose.  
 CC -!- SUBUNIT: THE RESULTING SUCRASE AND ISOMALTASE SUBUNITS STAY  
 CC ASSOCIATED WITH ONE ANOTHER IN A COMPLEX BY NON-COVALENT LINKAGES.  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Brush border.  
 CC -!- PTM: THE PRECURSOR IS PROTEOLYTICALLY CLEAVED WHEN EXPOSED TO  
 CC PANCREATIC PROTEASES IN THE INTESTINAL LUMEN.  
 CC -!- PTM: SULFATED (BY SIMILARITY).  
 CC -!- DISEASE: Defects in SI are the cause of disaccharide intolerance  
 CC I (MIM:222900).  
 CC -!- MISCELLANEOUS: THERE IS A HIGH DEGREE OF HOMOLOGY BETWEEN THE  
 CC ISOMALTASE AND SUCRASE PORTIONS (41 % OF AMINO ACID IDENTITY)  
 CC INDICATING THAT THIS PROTEIN IS EVOLVED BY PARTIAL GENE  
 CC DUPLICATION.  
 CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.  
 CC -!- SIMILARITY: Contains 1 P-type (trefoil) domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 CC EMBL; X63597; CAA45140.1; -;  
 CC EMBL; M22616; AAA60551.1; ALT\_SEQ.  
 CC PIR; S36082; UAHU.  
 CC Genew; HGNC:10856; SI.  
 CC MIM; 222900; -;  
 CC GO; GO:0005903; C:brush border; TAS.  
 CC GO; GO:0005794; C:Golgi apparatus; TAS.  
 CC InterPro; IPR000322; Glyco\_hydro\_31.  
 CC InterPro; IPR000519; P\_trefoil.  
 CC Pfam; PF01055; Glyco\_hydro\_31; 2.  
 CC Pfam; PF00088; trefoil; 2.  
 CC SMART; SM00018; PD; 2.  
 CC PROSITE; PS00025; P\_TREFOIL; 1.  
 CC PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 2.  
 CC PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.  
 CC Multifunctional enzyme; Transmembrane; Glycoprotein; Hydrolase;  
 CC Glycosidase; Repeat; Signal-anchor; Sulfation; Disease mutation.  
 CC INIT\_MET 0 0 BY SIMILARITY.  
 CC CHAIN 1 1826 SUCRASE-ISOMALTASE, INTESTINAL.  
 CC CHAIN 1 1006 ISOMALTASE.  
 CC CHAIN 1007 1826 SUCRASE.  
 CC DOMAIN 1 11 CYTOPLASMIC.  
 CC DOMAIN 32 1826 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).  
 CC DOMAIN 42 59 LUMENAL.  
 CC DOMAIN 61 108 SER/THR-RICH.  
 CC DOMAIN 109 1006 P-TYPE.  
 CC DOMAIN 1007 1826 ISOMALTASE.  
 CC ACT\_SITE 504 504 SUCRASE.  
 CC ACT\_SITE 1393 1393  
 CC DISULFID 62 93 BY SIMILARITY.  
 CC DISULFID 76 92 BY SIMILARITY.  
 CC MOD\_RES 236 236 SULFATION (POTENTIAL).  
 CC MOD\_RES 238 238 SULFATION (POTENTIAL).  
 CC MOD\_RES 390 390 SULFATION (POTENTIAL).  
 CC MOD\_RES 399 399 SULFATION (POTENTIAL).  
 CC MOD\_RES 666 666 SULFATION (POTENTIAL).  
 CC MOD\_RES 762 762 SULFATION (POTENTIAL).  
 CC MOD\_RES 764 764 SULFATION (POTENTIAL).  
 CC CARBOHYD 98 98 N-LINKED (GLCNAC...)  
 CC CARBOHYD 436 436 N-LINKED (GLCNAC...)  
 CC CARBOHYD 454 454 N-LINKED (GLCNAC...)  
 CC CARBOHYD 660 660 N-LINKED (GLCNAC...)  
 CC CARBOHYD 903 903 N-LINKED (GLCNAC...)  
 CC CARBOHYD 1234 1234 N-LINKED (GLCNAC...)  
 CC CARBOHYD 1302 1302 N-LINKED (GLCNAC...)  
 CC CARBOHYD 1339 1339 N-LINKED (GLCNAC...)  
 CC CARBOHYD 1353 1353 N-LINKED (GLCNAC...)  
 CC CARBOHYD 1402 1402 N-LINKED (GLCNAC...)  
 CC CARBOHYD 1747 1747 N-LINKED (GLCNAC...)  
 CC CARBOHYD 1762 1762 N-LINKED (GLCNAC...)  
 CC VARIANT 1097 1097 Q -> P (in disaccharide intolerance I;  
 CC exhibits intracellular accumulation of  
 CC mannose-rich SI in the Golgi).  
 CC /FTID=VAR\_007854.  
 CC FT CONFLICT 661 661 MISSING (IN REF. 2).  
 CC SEQUENCE 1826 AA; 209272 MW; 3P7B4B66DFC9C8E CRC64;  
 Query Match 28.7%; Score 1329.5; DB 1; Length 1826;  
 Best Local Similarity 36.4%; Pred. No. 2.3e-87;  
 Matches 307; Conservative 144; Mismatches 309; Indels 83; Gaps 23;  
 QY 59 DVORLAVYASLEDSRLRVRITDADHPRWEVPODIIPR---PAPGDVLDHPASSAPLQ 115  
 DB 141 DINSVLFTTQNPFRFKITDPNNRREYVPHQYKVEFTGPTVSDLYD----- 190  
 QY 116 GRVLSAGSOLVLTVHASPPRFTVSRRTSGDTLFDTRAPG-LVFRDKYLETSALPAGRAS 174

```
Db 191 -----VKVQNPFSTQVIRKSNKGLFTDSIGPLVTSQYQIGSARLPSD--Y 236
Qy 175 LYGLGHTKSSFLRHDSFTLW--NADIGASYVDVNLVYGHSPYDMVR-APGTAHGV 230
Db 237 IYVIGSEVHK--RFRHLSWKTWPIFTRDQLPCDNNNLYGHQTFFMCIEDTSGRSPGV 294
Qy 231 LLSNGMDV-LYGSVVTYKTVGGVLDFFYFAGPPLAVDDQYTOLIAAPAMPYWSGF 289
Db 295 LMNSNAMEIFIOPTPIVTVRVGGILODFYLLIGDTPQVQVQYQQLVGLPAMPAYWNLGF 354
Qy 290 HOCRYGLVNSDLERVAVYAKARIPLEVMWTDIDYMDGKDFPTLDRVNFATAELRPVD 349
Db 355 QLSRWNYKSLDVVKEVVRNREAGIPFDQVTDIDYMDKXDFYDQVAFNG--LPQFVQ 412
Qy 350 RLHRNAQKVLILDPQIRVD--PIDATYGTFRVGMQOIFLKN--GNFVGNVWPGDVY 405
Db 413 DLHDHGQKTVIILDPAISICRRANGTYYATVERGNTQHVWINESDGSPILIGEVWPLTV 472
Qy 406 FPEVHPAAAEFAWEISIFRRTIPVDGLWIDMNEISNFYNPEP---MNALDDPPYIN 461
Db 473 YPDTNPNCIDMWNBCSIFHOEVQVYDGLWIDMNEVSSFIQSTGKCNVKNLYPPF--T 530
Qy 462 NDTGTRPINKTVRPLAVHYGGVTEYEEHNLFLGLAARATGRVLR-DTGRPFVLSRST 520
Db 531 POILDKLWYKSKICMDAVQNWG-KQYDVHSLYGYMAIATEAQVQKVPNKESFILTRST 589
Qy 521 FVSGRYTAYTCDNAATWGDRLYSINTMLSGLFQMPMGADICGFNGNTTEELCGRWI 580
Db 590 FAGSGRHAARHLGDNATASBEQWMSITGLMFLSFLGVLVGDADICGFAETTEELCRWM 649
Qy 581 QLCAPYFPRDHS-----IFTVRRELYLWPSVAASGRKALGLRYQLLPVFTLMVE 632
Db 650 QLCAPYFPRNENSDYEHQDPAFFQNSL-----LVKSRQYLIRYLLPFLITLFFK 704
Qy 633 AHMTGAPIARPLFFSYPHDVATYVDRQFLRGVLVSPVLPPTTVDAYPPAGRWYRL 692
Db 705 AHVFGETVAPVLPHEFYEDTNSWIEDTEFLWGPALLITPVLKQGDATVSAYIPDAIW-- 761
Qy 693 YDY-SLAVATRTGKVRPAPADTVNVHLTGTTLPLOQSALTTSRARTAFHLVALAE 751
Db 762 DYESGAKPWRKQVRDMYLPADKGLHURGGYIPIQEPDVTVTASRNPIGLVALGE 821
Qy 752 DGTASGLFLDGDSPYGRSDMSVRENYKIPNNKGAIKVKSEVNSVAQSRTLVIS 811
Db 822 NNTAKGDFWDDGETKTIONGY--ILYTFVSNN-----TLDIVCTHSSYQEGTTLAQ 875
Qy 812 KVVLMGHRSPAPKPLTVHVNAGAEVEASSAGTRYON--AGSLGGVAHIGGLSVVGSEF 869
Db 876 TVRKLG-----LTDSTVTRVRAENNOPWNAHSNFTYDASNOVLLIADLKLNGNF 926
Qy 870 ELK 872
Db 927 SVQ 929

RESULT 13
MGA HUMAN
ID MGA HUMAN STANDARD; PRT: 1856 AA.
AC O43451,
DT 30-MAY-2000 (Rel. 39, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Maltase-glucoamylase, intestinal [Includes: Maltase (EC 3.2.1.20)
DE (Alpha-glucoamylase), Glucoamylase (EC 3.2.1.3) (Glucan 1,4-alpha-
DE glucosidase)].
GN MGA OR MGA OR MGAML.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
```

```
RC TISSUE=Small intestine;
RA MEDLINE=98112863; PubMed=9446624;
RA Nichols B.L., Eldering J.A., Avery S.E., Hahn D., Quaroni A.,
RA Sterchi E.E.;
RT "Human small intestinal maltase-glucoamylase cDNA cloning. Homology
RT to sucrose-isomaltase";
RL J. Biol. Chem. 273:3076-3081 (1998).
RN [2]
RP REVISIONS TO 776; 1049; 1100; 1541; 1612 AND 1811.
RA Nichols B.L., Eldering J.A., Avery S.E., Hahn D., Quaroni A.,
RA Sterchi E.E.;
RP Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP CHARACTERIZATION.
RC TISSUE=Small intestine mucosa;
RA MEDLINE=89066802; PubMed=3143729;
RA Naim H.Y., Sterchi E.E., Lentze M.J.;
RT "Structure, biosynthesis, and glycosylation of human small intestinal
RT maltase-glucoamylase";
RL J. Biol. Chem. 263:19709-19717 (1988).
RN [4]
RP SULFATION.
RC MEDLINE=88082658; PubMed=3121301;
RA Danielssen E.M.;
RT "Tyrosine sulfation, a post-translational modification of microvillar
RT enzymes in the small intestinal enterocyte.";
RL EMO J. 6:2891-2896 (1987).
CC -1- FUNCTION: MAY SERVE AS AN ALTERNATE PATHWAY FOR STARCH DIGESTION
CC WHEN LUMINAL ALPHA-AMYLASE ACTIVITY IS REDUCED BECAUSE OF
CC IMMATURITY OR MALNUTRITION. MAY PLAY A UNIQUE ROLE IN THE
CC DIGESTION OF MALTED DIETARY OLIGOSACCHARIDES USED IN FOOD
CC MANUFACTURING.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-
CC linked D-glucose residues with release of D-glucose.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -1- SUBUNIT: Monomer.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Brush border.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN SMALL INTESTINE, GRANULOCYTE, AND
CC KIDNEY BUT NOT IN SALIVARY GLAND OR PANCREAS.
CC -1- PTM: N- AND O-GLYCOSYLATED.
CC -1- PTM: DOES NOT UNDERGO INTRACELLULAR OR EXTRACELLULAR PROTEOLYTIC
CC CLEAVAGE.
CC -1- PTM: SULFATED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.
CC -1- SIMILARITY: Contains 2 P-type (trefoil) domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF016833; AAC39568.2; -.
CC Genew: HGNC:7043; MGAM.
CC MN: 154360; -.
CC GO: GO:0005983; P:starch catabolism; TAS.
CC InterPro: IPR000322; Glyco_hydro_31.
CC InterPro: IPR000519; P_trefoil.
CC Pfam: PF01055; Glyco_hydro_31; 2.
CC Pfam: PF00088; trefoil; 2.
CC SMART: SM00018; PD; 2.
CC PROSITE: PS00129; GLYCOSYL HYDROL_F31_1; 2.
CC PROSITE: PS00025; P_TREFOIL; 1.
CC Multifunctional enzyme; Transmembrane; Glycoprotein; Hydrolase;
CC Glucosidase; Repeat; Signal-anchor; Sulfation.
CC INIT MET 0 0
CC DOMAIN 1 12 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 13 33 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
CC (POTENTIAL).
```

FT	DOMAIN	34	1856	LUMENAL (POTENTIAL) .	
FT	DOMAIN	37	83	SER/THR-RICH.	
FT	DOMAIN	88	132	P-TYPE 1.	
FT	DOMAIN	197	914	MALTASE.	
FT	DOMAIN	952	998	P-TYPE 2.	
FT	DOMAIN	1066	1812	GLUCOAMYLASE.	
FT	ACT_SITE	528	528	BY SIMILARITY.	
FT	ACT_SITE	1419	1419	BY SIMILARITY.	
FT	DISULFID	89	117	BY SIMILARITY.	
FT	DISULFID	100	116	BY SIMILARITY.	
FT	DISULFID	111	129	BY SIMILARITY.	
FT	DISULFID	967	982	BY SIMILARITY.	
FT	DISULFID	977	995	BY SIMILARITY.	
FT	MOD_RES	415	415	SULFATION (POTENTIAL) .	
FT	MOD_RES	424	424	SULFATION (POTENTIAL) .	
FT	MOD_RES	1281	1281	SULFATION (POTENTIAL) .	
FT	CARBOHYD	134	134	N-LINKED (GLCNAC. .)	(POTENTIAL) .
FT	CARBOHYD	294	294	N-LINKED (GLCNAC. .)	(POTENTIAL) .
FT	CARBOHYD	456	456	N-LINKED (GLCNAC. .)	(POTENTIAL) .
FT	CARBOHYD	457	457	N-LINKED (GLCNAC. .)	(POTENTIAL) .
FT	CARBOHYD	478	478	N-LINKED (GLCNAC. .)	(POTENTIAL) .
FT	CARBOHYD	706	706	N-LINKED (GLCNAC. .)	(POTENTIAL) .
FT	CARBOHYD	748	748	N-LINKED (GLCNAC. .)	(POTENTIAL) .
FT	CARBOHYD	826	826	N-LINKED (GLCNAC. .)	(POTENTIAL) .
FT	CARBOHYD	884	884	N-LINKED (GLCNAC. .)	(POTENTIAL) .
FT	CARBOHYD	911	911	N-LINKED (GLCNAC. .)	(POTENTIAL) .
FT	CARBOHYD	976	976	N-LINKED (GLCNAC. .)	(POTENTIAL) .
FT	CARBOHYD	988	988	N-LINKED (GLCNAC. .)	(POTENTIAL) .
FT	CARBOHYD	1254	1254	N-LINKED (GLCNAC. .)	(POTENTIAL) .
FT	CARBOHYD	1322	1322	N-LINKED (GLCNAC. .)	(POTENTIAL) .
FT	CARBOHYD	1363	1363	N-LINKED (GLCNAC. .)	(POTENTIAL) .
FT	CARBOHYD	1387	1387	N-LINKED (GLCNAC. .)	(POTENTIAL) .
FT	CARBOHYD	1602	1602	N-LINKED (GLCNAC. .)	(POTENTIAL) .
FT	CARBOHYD	1671	1671	N-LINKED (GLCNAC. .)	(POTENTIAL) .
FT	CARBOHYD	1841	1841	N-LINKED (GLCNAC. .)	(POTENTIAL) .
FT	CARBOHYD	1846	1846	N-LINKED (GLCNAC. .)	(POTENTIAL) .
FT	CARBOHYD	1856	1856	N-LINKED (GLCNAC. .)	(POTENTIAL) .
SQ	SEQUENCE	1856	AA; 209721	NM; 81E7AA0CABAD07D	CRC64;

Query Match 28.3%; Score 1310.5; DB 1; Length 1856;

Best Local Similarity 36.28; Pred No. 5.5e-96;

Matches 307; Conservative 143; Mismatches 310; Indels 87; Gaps 26;

QY	59	DVORLAVYASLETDSRLVRITDADPRWEVPODIIIPRAPGDLVHDAPPASAPLQGRV	118
DB	166	NVDNVLTAQVTSNRPFKLTDQNNRFVPEHVQS-----FSGNAAASLTQ---	215
QY	119	LSFAGSDLVTVASPRFTVSRRSTGDTLPDTPAG-LVPRDKYLEVTSALPAGRSLVG	177
DB	216	-----VEISGRQPFISIKVTRSNRNVLFDSISIGPLLPADQLQLSTRLPS--TNVYG	264
QY	178	LGEHTKSSRLRHNDSTFLW---NADIGASVDVNLGSHPPFMDVR-APGTAHVLLLS	233
DB	265	LGEHVHQY--RHDMMKWTWPIFNRTTPNGNTNLYGATTFCLCLEDASGLSFGVFLMN	322
QY	234	SNQHD-VLYGGSVYTKVIGVLDVFFAGNPFLAVVDQYTLIARPAWPMYMSFGHQC	292
DB	323	SNAMVVLQAPAITVRTIGGILDVYFLGNTPEQVQVQVYLEIGRPAFPSYWALGFHLS	382
QY	293	RYGVLNVSLERVARAKARIPLEVMWTDIDYMDGFKDTLDVNFETAELRPFVDRLK	352
DB	383	RYEYGLDNNREVVNRRAQLPYDQHADIDYMDERRDFTYDSVDFKG--FPEFYNELH	440
QY	353	RNAQKVLIDPGIRVDPIDA-TYGTVPVGMQODIFLKRNG--TNFVGNVPGDGVFPDF	409
DB	441	NNGQKLVIVDPALSNSSSKYPGPDYDRGDMKIWNSSDGYTPPLIGVWPGQTFFPY	500
QY	410	MHPAAEFWAREISLFRRTIPVGLMTDMNEISFNYPE----PMNALDPPYRINNDGT	465
DB	501	TNPNCAVWTKFELFHNQVEFGIWMIDMNEVSNFVDSVSGCSTNNLNPPF-----T	554
QY	466	GRPLNN---KTVPLAVHVGGVTEYEHLNGLLEARATGRGVLR-DTGRRFVLSRST	520
DB	555	PRILDGVLFCMLDQVHMG-KQYDIHNLGYSMATAAEAAKTVFPNKRSLFTRST	613

QY	521	FVSGRYTAYWTGDAATGWDLRYISINTMLSPGLFGPMIGADICGNGNTTEELCGRWI	580
DB	614	FAGSGKFAAHLWDNTATWDDLRSIIPGVLEFNLFGIPMWGPDICGFALOTPEELCRMM	673
QY	581	QLGAFYFSPSRDHS-----AJFTVRRELZYLWPSVAASGRKALGLRYOLLPVFYTLMY	631
DB	674	QLGAFYFSPSRNNGQYKQDPPASFGADSL-----LNSRHYLNIRYTLPLVLYTLFF	727
QY	632	EAHMTGAPIARPLFSYPHDVATGVDRQFLGLGVLSVPLRPGPTTYDAYFPAGRWYR	691
DB	728	RAHSRGDTVARPLHFEYEDNSTWVHQFGLGPGELLITPVLDGAEBKWAYVDAVM--	785
QY	692	LYDYSLAVATGTGH-VRLPAPADTVNVHLTGTLPLQOSALTTSRARTAFHLLVALA	750
DB	786	-YDVTSGQVWRKQVEMELPGDKIGLHLGGVIFPTQDNTTLLASRKNPLGLIILD	844
QY	751	EDGTASGYFLDDGDSPEYGRSDMSVMRFNYKIPNNKGAIKVKSVEVHNSYAQSRTLVI	810
DB	845	ENKEAGBLFWDDGETKDTVANKVYLLCEFS--VTQN---RLEVNISQSTYKDPNNLAF	898
QY	811	SKVVLMGHRSAPAKPLTVHNSAEVASSAGTRYONAGLGGVAVHIGLSLVGE---	867
DB	899	NBIKILGTEEPS---NVTVKGHVPSQTSPTV-TYDSNL-----KVAITIDILLGEAYT	950
QY	868	-EFELKV	873
DB	951	VEVSIKI	957

RESULT 14

ID	YAJ1 SCHPO	STANDARD;	PRT;	993	AA.
AC	Q09901;				
DT	01-FEB-1996	(Rel. 33, Created)			
DT	01-OCT-1996	(Rel. 34, Last sequence update)			
DT	28-FEB-2003	(Rel. 41, Last annotation update)			
DE	Putative family 31 glucosidase C30D11.01c precursor (EC 3.2.1.1.)				
GN	SPAC30D11.01C OR SPAC56F8.01.				
OS	Schizosaccharomyces pombe (Pission yeast).				
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;				
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;				
OX	NCBI_TaxID=4896;				
RN	{1}_TaxID=4896;				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=972;				
RX	MEDLINE=21848401; PubMed=11859360;				
RA	Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,				
RA	Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,				
RA	Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,				
RA	Collins M., Connor R., Cronin A., Davis P., Reltwell T., Fraser A.,				
RA	Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,				
RA	Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagele K.,				
RA	James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,				
RA	Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,				
RA	Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitch E.,				
RA	Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,				
RA	Skellton J., Simmonds M., Squares R., Squares S., Stevens K.,				
RA	Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,				
RA	Woodward J., Volkhardt G., Aert R., Robben J., Grymonprez B.,				
RA	Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,				
RA	Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,				
RA	Borzym K., Langer I., Beck A., Lehrsach H., Reinhardt R., Pohl T.M.,				
RA	Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,				
RA	Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,				
RA	Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,				
RA	Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,				
RA	Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,				
RA	Domínguez A., Revuelta J., Moreno S., Armstrong J., Forsburg S.L.,				
RA	Cerrutti L., Lowe T., McCombie W.K., Paulsen I., Potashkin J.,				
RA	Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;				
RT	*The genome sequence of Schizosaccharomyces pombe.*;				



RL Nature 415:871-880(2002).  
 CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: 267961; CAA91887.1; -  
 CC EMBL: 269728; CAA93572.1; -  
 CC PIR: T38598; T38598.  
 CC GeneDB\_Spombe: SPAC30D11.01c; --  
 CC InterPro: IPR000322; Glyco hydro 31.  
 CC Pfam: PF01055; Glyco hydro 31; 1-  
 CC PROSITE: PS00129; GLYCOSYL\_HYDROL\_F31\_1; FALSE\_NEG.  
 CC PROSITE: PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.  
 CC Hypothetical protein; Hydrolase; Glycosidase; Glycoprotein; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 993  
 FT  
 FT DOMAIN 24 36  
 FT CARBOHYD 7 7  
 FT CARBOHYD 44 44  
 FT CARBOHYD 89 89  
 FT CARBOHYD 121 121  
 FT CARBOHYD 138 138  
 FT CARBOHYD 161 161  
 FT CARBOHYD 169 169  
 FT CARBOHYD 232 232  
 FT CARBOHYD 361 361  
 FT CARBOHYD 386 386  
 FT CARBOHYD 393 393  
 FT CARBOHYD 423 423  
 FT CARBOHYD 447 447  
 FT CARBOHYD 480 480  
 FT CARBOHYD 488 488  
 FT CARBOHYD 545 545  
 FT CARBOHYD 548 548  
 FT CARBOHYD 614 614  
 FT CARBOHYD 673 673  
 FT CARBOHYD 814 814  
 FT CARBOHYD 826 826  
 FT CARBOHYD 835 835  
 FT CARBOHYD 846 846  
 FT CARBOHYD 910 910  
 FT CARBOHYD 940 940  
 FT CARBOHYD 987 987  
 FT SEQUENCE 993 AA; 111043 MW; EB9471F19AAA9BC9 CRC64;  
 Query Match 27.7%; Score 1282.5; DB 1; Length 993;  
 Best Local Similarity 32.9%; Pred. No. 2.4e-84;  
 Matches 286; Conservative 156; Mismatches 287; Indels 141; Gaps 20;  
 QY 59 DVORLAVASLETSRLVRITDADHPREVQDIIPREPAGDVLHDAPPASSAPLQGRV 118  
 DB 114 DYPYLLNLSYDETERVHISIDLNQTFQ-----LSNRDWDAPFYRS 159  
 QY 119 LSPAGS-DLVLTVAHSPFRFTVSRRTGTLDTPAG-LVPRDKYLEVTSALPAGRASLY :76  
 DB 160 SNFSGNLQVNFSTDPFFWITRIADDDQVLDTRGNPLIPDQVIELTNN-VEDYVNY 218  
 QY 177 GLCEHTKSPRLRHNSFTLWNAIDIGASVYDNLVGHSPFYMDVR-----APGTAH 227  
 DB 219 GL-SGQQSFRLGNLTKTFTATGYSDS-PEANVGSHPFYMEQRYIPIGTNTVTSASH 276  
 QY 228 GVLLLSNGMDVLVYGSYVTKYVIGGVLDYFPAGP--NPLAVDQYQTOLIARPAMPYVW 285  
 DB 277 GVLMSSNGMEVLLRSTYIKYRMIGGIIDLVYVSGSTVSPKTYIQYVQVSGTPTMQPYW 336  
 QY 286 SFGFHQCRGYLNVSLERLVARYAKARIPLEVMWTDIDYMDGFKDFTLDRVNFPAEALR 345

DB 337 SLGFQMSRWGYKTSIDLINMESYLNASNIPTEGFWDNDIDYMFSEFTFTVNSTAFPPQTL 396  
 QY 346 PVDRLRHNAQKYVLLDPGI-RVDP-----IDATYGTFRGMQDIFLKR-NGTNFVGNVW 400  
 DB 397 DFRSLDSESHQHYVPVLDPAIYAANPNKSAADRTYYPYSGFEDNIFIKPNNGSAYVGNW 456  
 QY 401 PGDVTYFPDMPAAAFWAREI-----SLFRTTIPVDGLWIMDRISNP----- 444  
 DB 457 PGFVYVPDTPAVLYWKQGIILNLSTAFGSNYGYDLFPFSGCLDNNEPTSCFICGCGSD 516  
 QY 445 -----YN-PEPMNA----- 452  
 DB 517 LUKNPVHPFSLPGDVKNKVSYPEDFNATNTTEYKSVSPASOSQYKATATSEKSHETP 576  
 QY 453 -----LDPPYRINDGTGRPINNKTVRPLAVHYGGVTEYEHNLFLGLEAR 499  
 DB 577 SSESILINGKPEFSINYPYALDITDTHDLAQFGVSPNATMHGNTLRYNLNTYGYSESK 636  
 QY 500 ATGRGVLR-DTGRRPVLSRSTFVSGRYTAYWGDNAATWGDLYRSINTWLSFGLFGMP 558  
 DB 637 ISFEALNSIQPNIRPFLSRSTFVSGRYAAHGLGDNKQMSDMVSSISSILTFLNLGIP 696  
 QY 559 MIGADICGFNGNTTEELCGRWIQLCAFYFPRDRHSAIFTVRRELYLWPSVAASGRKALGL 618  
 DB 697 MYGADVGVNGNTDEELCARWMLGAFLPFYRNHNSLGSIPQEPFRWASVAESAIEI 756  
 QY 619 RYOLLPYFTLYMEAHMTGAPIARLPFSYHDVATYVDQFLLGRGVLPVSPLEPGPT 678  
 DB 757 RYSLLPYVYTLMTASVDGTMPVRPFFEFKQISLASVDKQFMIGTALLISPALEPNT 816  
 QY 679 TVDAYFPAGR---WYRLYDLSAVATRTGKVRLPAPADTVNHLTGTTILPQOSALT 735  
 DB 817 YIQGIIPGNDITWDTNHSV-INHDYDENTSAPLGYNYIARVGGNIIPLOQPGYTT 875  
 QY 736 SPARKTAFHLLVALAEDGTAGCYLFDLDDGDEYGRSRSDMSVRENYKIPNKGAKYKS 795  
 DB 876 YESRNPYSLLIAMDNNGFASGSLYIDGISMQ-----TNSISLVKLNS 919  
 QY 796 EVVHNSYAQSRTLV-----ISKVVLNGHRSR 821  
 DB 920 NSNTITCVVSGTMVSSPSLANITILGLSNP 949  
 RESULT 15  
 ID AMYG DEBOC STANDARD; PRT; 958 AA.  
 AC P22861; Q92336;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-AUG-1991 (Rel. 19, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Glucoamylase 1 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)  
 DE (1,4-alpha-D-glucan glucohydrolase).  
 GN GAMI.  
 OS Debaryomyces occidentalis (Yeast) (Schwanniomyces occidentalis).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Debaryomyces.  
 CX NCBI\_TaxID=27300;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN=ATCC 26076;  
 RX MEDLINE=91071592; PubMed=1979298;  
 RA Dohmen R.J., Strasser A.W.M., Dahlems U.M., Hollenberg C.P.;  
 RT "Cloning of the Schwanniomyces occidentalis glucoamylase gene (GAMI)  
 RT and its expression in Saccharomyces cerevisiae.";  
 FL Gene 95:111-121(1990).  
 RN [2]  
 RP SIMILARITY TO OTHER FAMILY 31 MEMBERS.  
 RX MEDLINE=92077121; PubMed=1743281;  
 RA Naim H.Y., Niermann T., Kleinhaus U., Hollenberg C.P.,  
 RA Strasser A.W.M.;  
 RT "Striking structural and functional similarities suggest that  
 RT intestinal sucrase-isomaltase, human lysosomal alpha-glucosidase and



RT Schwannomyces occidentalis glucoamylase are derived from a common  
 RL ancestral gene";  
 CC FEBS Lett. 294:109-112(1991).  
 CC -1- FUNCTION: THIS GLUCOAMYLASE HAS A SPECIFICITY TOWARD BOTH  
 CC ALPHA-1,4 AND ALPHA-1,6 LINKAGES.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-  
 CC glucose residues successively from non-reducing ends of the chains  
 CC with release of beta-D-glucose.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M60207; AAA33923.1; --  
 CC PIR; JN0102; JN0102.  
 CC InterPro: IPR000322; Glyco hydro\_31.  
 CC Pfam; PF01055; Glyco hydro\_31; 1.  
 CC PROSITE; PS00129; GLYCOSYL HYDROL\_F31\_1; 1.  
 CC PROSITE; PS00707; GLYCOSYL HYDROL\_F31\_2; 1.  
 CC Hydrolase; Glycoelidase; Polysaccharide degradation; Glycoprotein;  
 CC Signal.  
 CC  
 CC 1 22 POTENTIAL.  
 CC CHAIN 23 958 GLUCOAMYLASE 1.  
 CC ACT\_SITE 470 470 BY SIMILARITY.  
 CC DOMAIN 26 41 SER-RICH.  
 CC FT DOMAIN 530 542 SER/THR-RICH.  
 CC FT CARBOHYD 61 61 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC FT CARBOHYD 78 78 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC FT CARBOHYD 107 107 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC FT CARBOHYD 197 197 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC FT CARBOHYD 403 403 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC FT CARBOHYD 416 416 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC FT CARBOHYD 513 513 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC FT CARBOHYD 580 580 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC FT CARBOHYD 602 602 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC FT CARBOHYD 813 813 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC FT CARBOHYD 907 907 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC SEQUENCE 958 AA; 106507 MW; 47938DB9BC308260 CRC64;  
 Query Match 27.6%; Score 1279.5; DB 1; Length 958;  
 Best Local Similarity 34.4%; Pred. No. 3.7e-84;  
 Matches 315; Conservative 137; Mismatches 313; Indels 151; Gaps 27;  
 30 AARTVLAVAVT-----MEGALRAEAAATGGRSGTGDVORLAVYASLETDSRLVRITDAD- 83  
 70 AAKGYDLVNTNTPRGTLTGILKEATNIYGYDFYLNLTAV--EYQADTRLNHVEPTDL 127  
 84 HPRVEVDQIIPRP-APGDVLHDAPPASSAPLOGRVLSPPAGSDLVLTVAHSPREFTVSR 142  
 128 SDVFLPEHLVVKPELVGD-----AQSYNPD-----NSDLVFEYSNTDFSEVIRS 173  
 143 STGDTLFDTPAG-LVFRDKYLEVTSALPAGRASLYGLGE-----HTKSSFRLEHNSFT 195  
 174 STKEVLFSTKGNPLVFSNQFIQFNSSLPKNHV-ITGLGESIHGIVNEPQSVK-----T 225  
 196 LWNADIGASYVDVNLVYKSHPPYMDVR-APGTAGVGLVLLSNGMDVLYGGSYVYKVGIV 254  
 226 LFANDVVG-DPIDGNIYGVHPVQLDQRYDTTTHAVYVWRTSAIQEVLIGESITWRLSGV 284  
 255 LDEYFFAGNPVLAVVDQVLTOLIAAPAMPYWSFGFHCRCYGLNVLNVDLERVVARYAKARI 314  
 285 IDLFFSGFTPKDAIQVYKIGIPLAQPYWSLGHQRCMGYDTIEKLSEVENFKAFNI 344  
 315 PLEVMTIDIDYMGFKDFTLDRVNFATAELRPFVDRLHRNAQKYVLLIDPGIRV-DFIDA 373  
 345 PLETIWSIDYMSYKDYDTPHPPFLDEYRKFLDELKKNQHVPVILDAAIYVPENNA 404  
 374 T---YGVTVRGMQQDIFLKR-NGTNFVGNWVPGDVYFPDFMHPAAAEFWAREISLFRTI 429

Db 405 TDNEYQPHYGNEDTVFLKNDPDSGLYIGAVWQVTLF-----SRFLSRKSHDMXVI 455  
 QY 430 -----PVDGLWIDMNEISNF-----YNP----- 447  
 Db 456 KDWYELTFPDGIWADMNEVSSFCVGSCTGKYFENPAYPPFTVGSKATSVPGVDVSNAS 515  
 QY 448 -----EPNAL-----DDPPYRINDGTGRPINNK 472  
 Db 516 EWKSIQSSISATKTSSTSVSSSSSTIDYNTLAPGKGNINIPYAIYNNQGSDDLATH 575  
 QY 473 TVRPLAVHYGVVTEYEENLFGLEABATRGVLR-DTGRPPFVLRSSTFGSGRYTAYW 531  
 Db 576 AVSPNATHADGTVEYDHNLYGYLQENATYHALLEVFNKEPPMISRSSTFPRACKWIGHW 635  
 QY 532 TGDNAATGDLRYISINTMLSGLFQPMWIGADICGFNGNTTEELCGRWIQLGATYPPSRD 591  
 Db 636 GGDNTADWAYAYFISIPQAFSMGIAGLPFGADVCGFNGNSDSELCRMMQLGSEFFPYRN 695  
 QY 592 HSAFTVRRELYLWPSVAASGRKALGRYQLLPYFTLMVEAHMTGAPIARPLFFSYPHD 651  
 Db 696 HNYLGAIDQEPYVWESVABATRTSMARLYLLPYTYLLHESHTTGLPILRAFWSQFPND 755  
 QY 652 VATYGVDRQFLLGRGVLSVPLEPGPTTVDAYFP-AGRMYRLYDY--SLAVATRTGKHVR 708  
 Db 756 RSLSGVDNQFFVGDLVVTPLVLEPGVDKVGVPFGAGKEEVYDWTQREVFHKDGKNET 815  
 QY 709 LPAPADTVNVHLTGTTLPLQOQSALTTSRAPRTAFHLLVALAEDGTAGYLLDDGDSPE 768  
 Db 816 LDAPLGHIPLHIRGNVLTPTQEPGYTVAESHQNPFGLLVALDNDGKAQGSLLDDGESLV 875  
 QY 769 YGRRSDWMSVRFNYKIPNNKGAIKVKSEVHNHNSYAQSRTLVIKVLMLGHRSPAAPKKLT 828  
 Db 876 VD-----SSLLVSFSVSDN-----TLSASPSGDYKADQPLANVTILGVGHK----PK-S 919  
 QY 829 VVNSAEVEASSSAGT 844  
 Db 920 VKFENANVDFTYKKST 935

Search completed: October 27, 2003, 10:20:20  
 Job time : 20.1941 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 ; Search time 22.9249 Seconds  
(without alignments)  
3678.964 Million cell updates/sec

Title: US-10-043-418-1  
Perfect score: 4630  
Sequence: 1 MATVGVLLCLCLFAPRL.....IGGLSLVVGEEFELKVMSY 877

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues  
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_76:\*

1: Pir1:\*

2: Pir2:\*

3: Pir3:\*

4: Pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4630	100.0	877	2 S65057	alpha-glucosidase
2	2451	52.9	902	2 T48531	alpha-glucosidase
3	2422.5	52.3	903	2 T09143	alpha-glucosidase
4	2370	51.2	913	2 JC5463	alpha-glucosidase
5	2100	45.4	915	2 H96709	hypothetical prote
6	1904	41.1	855	2 T47534	hypothetical prote
7	1470	31.7	864	2 JC4624	alpha-glucosidase
8	1396.5	30.2	952	1 A32609	alpha-glucosidase
9	1370	29.6	1827	1 A23945	sucrose alpha-gluc
10	1348.5	29.1	985	2 JC4217	alpha-glucosidase
11	1329.5	28.7	1827	1 U0HU	sucrose alpha-gluc
12	1302	28.1	995	2 T50267	probable family 31
13	1282.5	27.7	993	2 T38598	probable family 31
14	1279.5	27.6	958	1 JN0102	glucan 1,4-alpha-g
15	1279	27.6	1841	2 T10799	sucrose alpha-gluc
16	1194.5	25.8	719	2 JC1200	alpha-glucosidase
17	1152.5	25.8	1743	2 T15893	hypothetical prote
18	1142	24.7	1070	2 S19686	alpha-glucosidase
19	1113.5	24.0	919	2 T16693	hypothetical prote
20	1002	21.6	856	2 T22575	hypothetical prote
21	895	19.3	779	2 AE2402	alpha-glucosidase
22	853	18.4	910	2 T22050	hypothetical prote
23	853	18.4	924	2 T22044	hypothetical prote
24	826.5	17.9	919	2 T07391	probable alpha-glu
25	815	17.6	763	2 AG1460	alpha-glucosidase
26	813	17.6	941	2 T32449	hypothetical prote
27	801	17.3	763	2 AH1097	alpha-glucosidase
28	775.5	16.7	818	2 AC2412	alpha-glucosidase
29	762.5	16.5	954	2 S46105	glucan 1,4-alpha-g

ALIGNMENTS

RESULT 1

S65057  
alpha-glucosidase (EC 3.2.1.20) - barley  
C:Species: Hordeum vulgare (barley)  
C:Date: 20-Jul-1996 #sequence\_revision 27-Feb-1997 #text\_change 22-Jun-1999  
C:Accession: S65057; S65058  
R:Tibbot, B.K.; Skadsen, R.W.  
submitted to the EMBL Data Library, March 1995  
A:Description: Molecular cloning and characterization of a gibberellin-inducible alpha-glucosidase  
A:Reference number: S65057  
A:Accession: S65057  
A:Molecule type: mRNA  
A:Residues: 1-877 <TIB>  
A:Cross-references: EMBL:U22450; NID:G944900; PIDN:AAB02985.1; PID:G944901  
R:Tibbot, B.K.; Skadsen, R.W.  
Plant Mol. Biol. 30, 229-241, 1996  
A:Title: Molecular cloning and characterization of a gibberellin-inducible, putative alpha-glucosidase  
A:Reference number: S65058; MUID:96178863; PMID:8616248  
A:Accession: S65058  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 85-115,'S',117-121,123-170,'D',172-366,'G',368-529,'H',531-685,'G',687-766  
A:Cross-references: EMBL:U22450  
C:Superfamily: lysosomal alpha-glucosidase; sucrose/isomaltase homology; trefoil homo:  
C:Keywords: Glycosidase; hydrolase  
F:125-769/Domain: sucrose/isomaltase homology <SIM>

Query Match	100.0%	Score 4630;	DB 2;	Length 877;
Best Local Similarity	100.0%;	Pred. No. 0;	Mismatches	0;
Matches	877;	Conservative	0;	Indels 0; Gaps 0;
QY	1	MATVGVLLCLCLFAPRLCSSKEEGLAARTVLAVATVMEGALRAEAATGGRSSTGDV	60	
DB	1	MATVGVLLCLCLFAPRLCSSKEEGLAARTVLAVATVMEGALRAEAATGGRSSTGDV	60	
QY	61	QLAVYASLETDLSRLVRITADHPRNEVPDIIIPRAPGDLVDHPAPASSAPLQGRVLS	120	
DB	61	QLAVYASLETDLSRLVRITADHPRNEVPDIIIPRAPGDLVDHPAPASSAPLQGRVLS	120	
QY	121	PAGSLVLTVAHSPRFTVSRRTSGDTLFDTPAGLVFRDKYLEVTLSALPAGRSLYLIGE	180	
DB	121	PAGSLVLTVAHSPRFTVSRRTSGDTLFDTPAGLVFRDKYLEVTLSALPAGRSLYLIGE	180	
QY	181	HTKSSFRLRHNDSTFLMNADIGASVVDVNLXGSHPFYMDVPARGTAHCVLLSSNGMDVL	240	
DB	181	HTKSSFRLRHNDSTFLMNADIGASVVDVNLXGSHPFYMDVPARGTAHCVLLSSNGMDVL	240	
QY	241	YGGSVYTVKTVGGVLDFFYFAGPNPLAVVDQYTOIARPAHPYWSFGFHCRGYGLNVS	300	
DB	241	YGGSVYTVKTVGGVLDFFYFAGPNPLAVVDQYTOIARPAHPYWSFGFHCRGYGLNVS	300	
QY	301	DLERVVARYAKARIPLEVMTDIDYMDGFKDFTLDRVNFATAELRPFVDRLRNAQKTVL	360	

```
Db 301 DLERVVARYAKARIPLEVMWMTDIDYMDGFKDFTLDRVNFTEAELRPFVDRLHRNAQYVL 360
Qy 361 ILDPGIRVDPIDATYGTFRGQQQDIFLKRNGTNFVGNVWPGDVYPPDFMHPAAAEPWAR 420
Db 361 ILDPGIRVDPIDATYGTFRGQQQDIFLKRNGTNFVGNVWPGDVYPPDFMHPAAAEPWAR 420
Qy 421 EISLFRRTIPVDGLMIDMNEISNFPNPPMNAIDPPYRIINNDGTGRPINNKTVRPLAVH 480
Db 421 EISLFRRTIPVDGLMIDMNEISNFPNPPMNAIDPPYRIINNDGTGRPINNKTVRPLAVH 480
Qy 481 YGVVTEYEHNLFGLEARATGRGVLRDTGRPPFVLSRSTTVGSGRYTAYWTGDNATWG 540
Db 481 YGVVTEYEHNLFGLEARATGRGVLRDTGRPPFVLSRSTTVGSGRYTAYWTGDNATWG 540
Qy 541 DLRYNSINTMLSGFGMPMIGADICGFGNGNTTEELCGRWIOLGAFYPPSRDHSALFTVRR 600
Db 541 DLRYNSINTMLSGFGMPMIGADICGFGNGNTTEELCGRWIOLGAFYPPSRDHSALFTVRR 600
Qy 601 ELYLWPSVAASGRKALGURYQLLPYFTLTYEAHMTGAPIARPLFFSYPHDVATYGVDRQ 660
Db 601 ELYLWPSVAASGRKALGURYQLLPYFTLTYEAHMTGAPIARPLFFSYPHDVATYGVDRQ 660
Qy 661 FLRGRVLSPVLEPGPTVDAYFPAGRWYRLDYSLAVATRTGKVLTPAPADTVNHL 720
Db 661 FLRGRVLSPVLEPGPTVDAYFPAGRWYRLDYSLAVATRTGKVLTPAPADTVNHL 720
Qy 721 TGGTILPQOQSALTTSRARTAFHLLVALAEDGTASGLFLDDGDSPEYGRSDMSVRF 780
Db 721 TGGTILPQOQSALTTSRARTAFHLLVALAEDGTASGLFLDDGDSPEYGRSDMSVRF 780
Qy 781 NTKIPNNKGAIKYSEVWNSAQRTLVISKVLMGHRSPAAPKCLTVHNSAEVEASS 840
Db 781 NTKIPNNKGAIKYSEVWNSAQRTLVISKVLMGHRSPAAPKCLTVHNSAEVEASS 840
Qy 841 SAGTRYONAGGLGGVAHIGGLSLWGEPELKVMSY 877
Db 841 SAGTRYONAGGLGGVAHIGGLSLWGEPELKVMSY 877

RESULT 2
T48531
alpha-glucosidase 1 - Arabidopsis thaliana
N:Alternate names: protein T22p22, 110
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 21-Jul-2000
C:Accession: T48531
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft,
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24490
A:Accession: T48531
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-902 <BEV>
A:Cross-references: EMBL:AL163814
A:Experimental source: cultivar Columbia; BAC clone T22p22
C:Genetics:
A:Map position: 5
A:Map position: 78/2; 313/1; 390/1; 605/3; 747/1
A:Note: T22p22.110
C:Superfamily: lysosomal alpha-glucosidase; sucrase/isomaltase homology; trefoil homology

Query Match 52.9%; Score 2451; DB 2: Length 902;
Best Local Similarity 55.2%; Pred. No. 3e-178;
Matches 471; Conservative 136; Mismatches 191; Indels 56; Gaps 13;
Qy 59 DVORLAVIASLETDSRLVRITDADHPRWEVPQDIIIPRAPGVLDHAPPASSAPLQG-- 116
Db 69 DIKSLNLHVSLETSLRIRITDSSQORWEIPETVIPRAG----NHSRRFSTEEDGDN 123
Qy 117 ----RVLSAGSDLVLTVH-ASFRFTVSRRTGDTLFTAFG-----LVFRKYLEV 164
Db 124 SPENNFLADPSSDLVFTLHNTTTPFGFSVSRSSGDLFDTSPPSDSDNTYTFIKDQFLQL 183
```

```
Qy 165 TSALPAGRASLYGLGEHTKSSFRRLRHNDSTFLMNADIGASYVDVNLVYSGHPPFYMDVRA-- 222
Db 184 SSALPENASNLVIGEGHTKRSFRLLIPGETMTLMNADIGSENPVNLVYSGHPPFYMDVRGSK 243
Qy 223 ----PGTAGVULLSSNGMDVLYGGSYYTYKIVGVLDFFYFAGNPPLAVVDQVLTOLIA 278
Db 244 GNEBAGTTHGVLLNSNGMDVKYEGHRTYNNVIGVIDLYYFAGSPSPMVMNQVTEIGR 303
Qy 279 PAPMPVWSEFGHOCRYGVLNVSDLRVVARAKARIPLVNVMTDIDYMDGFKDFTLDRVN 338
Db 304 PAPMPVWSEFGHOCRYGVLNVSDLRVVARAKARIPLVNVMTDIDYMDGFKDFTLDRVN 363
Qy 339 FTAAELRPVDFLHRNAQYVLIIDPGIRVDPIDATYGTFRGQQQDIFLKRNGTNFVGN 398
Db 364 FPDKMQSPVDTLKXGQKYVLIIDPGIGV--DSSGYTYNRGMEADVFIKRNCEPYLGE 420
Qy 399 VMPGVDPVDFPMHFAAEPFAREISLFFRTIPVGLWIDMNEISNFPNPEM--NALDD 455
Db 421 VMPGVDPVDFPMHFAAEPFAREISLFFRTIPVGLWIDMNEISNFPNPEM--NALDD 455
Qy 456 PPYRINNDGTGRPINNKTVRPLAVHYGVVTEYEEHNLFGLEARATGRGVLRDTGRPPFV 515
Db 480 PPYRINNDGTGRPINNKTVRPLAVHYGVVTEYEEHNLFGLEARATGRGVLRDTGRPPFV 515
Qy 516 LSRSTFVSGRYTAYWTGDNATWGDLRYISINTMLSGFGMPMIGADICGFGNGNTTEEL 575
Db 540 LSRSTFVSGRYTAYWTGDNATWGDLRYISINTMLSGFGMPMIGADICGFGNGNTTEEL 575
Qy 576 CQRWTLQGFYFPFSDHSAIFTVRBELVLPWSVAASGRKALGRLYQLLPYFTLTYEAHM 635
Db 600 CQRWTLQGFYFPFSDHSAIFTVRBELVLPWSVAASGRKALGRLYQLLPYFTLTYEAHM 635
Qy 636 TGAPIARPLFFSYPHDVATYGVDRQFLRGRVLSVPLEPGPTVDAYFPAGRWYRLDY 695
Db 660 SGNPIARPLFFSPQDTKTIEIDSOFLGKSIWSPALKQGAVALDAYFPAGRWYRLDY 719
Qy 696 SLAVATRTGKVLTPAPADTVNHLTGGTILPQOQSALTTSRARTAFHLLVALAEDGTA 755
Db 720 SLAVATRTGKVLTPAPADTVNHLTGGTILPQOQSALTTSRARTAFHLLVALAEDGTA 755
Qy 756 SGLYPLDDGDSPEYGR---RSDMSVVRNFKYKIPNNKGAIKYSEVWNSAQRTLVISK 812
Db 780 SGLYPLDDGDSPEYGR---RSDMSVVRNFKYKIPNNKGAIKYSEVWNSAQRTLVISK 812
Qy 813 VTLMGHRSPAAPKCLTVHNSAE-----VEASSAGTRYONAGGLGGVAHIGGLS 862
Db 838 VTLMGHRSPAAPKCLTVHNSAE-----VEASSAGTRYONAGGLGGVAHIGGLS 862
Qy 863 LVVGEFEFLKVAMS 876
Db 889 LVVGEFEFLKVAMS 902

RESULT 3
T09143
alpha-glucosidase (EC 3.2.1.20) - spinach
C:Species: Spinacia oleracea (spinach)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: T09143
R:Sugimoto, M.; Furui, S.; Suzuki, Y.
Plant Mol. Biol. 33, 765-768, 1997
A:Title: Molecular cloning and characterization of a cDNA encoding alpha-glucosidase
A:Reference number: Z16585; MUID:97238484; PMID:9132069
A:Accession: T09143
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-903 <SUG>
A:Cross-references: EMBL:D86624; NID:g2081626; PIDN:BAAL19924.1; PID:g2081627
A:Experimental source: strain Daeh
C:Superfamily: lysosomal alpha-glucosidase; sucrase/isomaltase homology; trefoil homo
C:Keywords: Glycosidase; hydrolase
F:145-799/Domain: sucrase/isomaltase homology <SIN>
```



```

QY 686 AGRVRLVDYSLAVATRTGKHVRLPAPADTVNVHLTGCTILPQQSALTTSRARRTAFHL 745
DB 720 RGNVYSLNYSVSAGTVSLSAPPDHTNVHHEGNIIVAMGEAMTTQAAARSTPFHL 779
QY 746 LVALAEDGTASGYLFLDDGDSPEY-GRSDMSWVRENYKIPNNKGAIKVSEVHNSYAO 804
DB 780 LVVMSDRVASTGEFLDNGIEWDGGPGKWTLVRFRAESINN--ITISEVVNRGYAM 837
QY 805 SRTLIVSKVLMGHRSPAPKLTIVHNSAEVASS-SAGTRYQNAQGLGGVAHIGLSL 863
DB 838 SQRWMDKITILGLKRRVKIKEYTVQKDAKAIKVKGLGRRRTSSHQGF-FVSVISDLRQ 896
QY 864 VVGEFEFLKV 873
DB 897 LVGQAFKLEL 906

RESULT 5
H96709
hypothetical protein F24J5.20 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: H96709
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
A:Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lucos, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: H96709
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-915 <STO>
A:Cross-references: GB:AB005173; NID:g5734722; PIDN:AAD49987.1; GSPDB:GN00141
C:Genetics:
A:Gene: F24J5.20
A:Map position: 1

Query Match 45.4%; Score 2100; DB 2; Length 915;
Best Local Similarity 46.6%; Pred. No. 1.7e-151;
Matches 431; Conservative 145; Mismatches 267; Indels 82; Gaps 18;

QY 7 LLLCLCLCLFAP-----RLCSKEEGPLAARTVLAVAVTMEGAL--RAEAATGG 53
DB 14 LLLALLILC-FSPQTQSYKTIGKYLVSII-EESP-----DGGFTGYLVQVQKN 58
QY 54 RSSTGDVORLAVASLEFDSRLVRITDADHPREVPQDIIIPRAPGDVLHDAPASAP 113
DB 59 KIYSGDITTLFLVFKHETDSRLVRHITDAKQORMEVPYNLLPRSQPPQGVKVIQKSRSP 118
QY 114 LQGRVLSPAGSDVLTVHASPFRFTVSRKSTGDTLFTAPGLVFRDKYLEVTSALPAGRA 173
DB 119 ITVQIS--GSELIYSYTDPTFAVKRSNHELTNTSSLVFKDQVLEISTSLPK-EA 175
QY 174 SLYLGHEHTKSS-FRLRHNDSTLWADIGASYVDVNLVYGHFFYFMDVRAFG--TAHGV 229
DB 176 SLYGLGENSQANGIKLVNEPYTLVTEDEVSAINLNTDLYGSHPMYMDLRNVGGKAYAHAV 235
QY 230 LLLSSNGMDVLYGGSYVYKVIIGGVLDFFYFAGNPLAVNTOYTOLINAPMPYWSFGF 289
DB 236 LLLSSNGMDVYFGSDLSYKVIIGGVLDFFYFAGNPLAVNTOYTOLINAPMPYWSLGF 295
QY 290 HOCRGYLNVSLERVARYAKARIPLEVMTDIDYMDGFKDFTLDRVNFETAELRPVD 349
DB 296 HOCRNGYHNLVSVEDVDVNYKAKIPLDVINWDDHMGHDKFTLNPVAYPRAKLLAFD 355
QY 350 RLHRNAQKYVLIDFGIRVDPIDATYGTFTVRGMOQDIFLKRNGTNFVGNWFGDVYFPDF 409

```

```

DB 356 KHKIGMKIIVINDPGIV--NASYGTFFORAAADVFIKYEKPFLLAQVWPGPYFPDF 412
QY 410 MHPAAAEFAWEISLFRRTIPVDGLWIDMNEISNFY-----PE----- 448
DB 413 LNPKTVMWGDEIKRFDLVPIDGLWIDMNEVSNFCSGLCTIPEGKQCPGSGPGWVCL 472
QY 449 -----PMALDPPPIRINDGTGRPNKTKVPLAVHYGGVTEYEERHNLFGLEEARATR 503
DB 473 DCKNITKTRWDDPPYKINATGVAVPVGFKTATSATHYNGVREYDAHSIYGFSETIATHK 532
QY 504 GYLRTDTRGPRFFVLSRSTFVSGRYTAYMTGDNAATWGLRYSINTMLSGFLGFMEMIGAD 563
DB 533 GLLVQGGKRPFLSRSTFVSGQYAAHWTGDNQGTQSLQVLSISIMLNFGLFGVPMVQSD 592
QY 564 ICGFNGNTTEELCGRWIOLGAFYPPSRDHSIAFTVRRRELYLWPSVAASGRKALGRLYQLL 623
DB 593 ICGFYPPQPTTEELCNWIEVGAFYPPSRDHANYSPQELYQWDTVADSARNALGMRYKIL 652
QY 624 PYPYTLMYEAMTGAIPARPLFFSYPHDVATVGVDRQPLGRLGVLSVPLEPGPTTVDAY 683
DB 653 PPLYTLNVEAMTGAIPARPLFFSYPHDVATVGVDRQPLGRLGVLSVPLEPGKTEVEAL 712
QY 684 FPAGRWYRLDYSLAVATRTGKHVRLPAPADTVNVHLTGCTILPQQSALTTSRARRTAF 743
DB 713 FPPGSMYEMFDMTQAVSKNGKRVTLPAFLNPNVHLVQNTILPTQOGLISKDARTTFF 772
QY 744 HLLVAL---AEGTASGYLFLDDGDSPEYGRSDMS--WVRFNYKIPNNKGAIKVSEVH 799
DB 773 SLVIAFPAGASGYATGKLYLDEDELPEMKLNGQSTYDFYASVGN--GTMKMMSQVKE 830
QY 800 NSVAGSRTILVSKVLMGHRSPAPKLTIVHVN--SAEVEASSSAGT-----RYQNAQ 850
DB 831 GKFALSKGVIRKSVLGRGAGQVSEIQINGSPMTKKIEVSKKHTYVIGLEDEENKS 890
QY 851 GLGGVAHIGGLSLVVGSEFELKVAM 875
DB 891 VM--VEVRGLEMLVKGDFNMSWKM 912

RESULT 6
T47534
hypothetical protein F16L2.150 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000
C:Accession: T47534
R:Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unseld, M.; Mewes, H.W.; Rudd, S
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24468
A:Accession: T47534
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-855 <JOR>
A:Cross-references: EMBL:AL162459
A:Experimental source: cultivar Columbia; BAC clone F16L2
C:Genetics:
A:Map position: 3
A:Introns: 69/3; 291/1; 427/3; 699/1
A:Note: F16L2.150
C:Superfamily: lysosomal alpha-glucosidase; sucrose/isomaltase homology; trefoil homo
Query Match 41.1%; Score 1904; DB 2; Length 855;
Best Local Similarity 43.8%; Pred. No. 1.3e-136;
Matches 404; Conservative 144; Mismatches 245; Indels 130; Gaps 21;

QY 1 MATVGVLLCLCLCLFAPRLCS-----SKEGPLAARTVLAVAVTMEGALRAEA 50
DB 1 MASCLLVAIILC-FSSLQCSNAILGKGYRLISMEKSPDGSFI-----GYLVQVQS 51
QY 51 TGRSSTGDVORLAVASLETSRLVRITDADHPREVPQDIIIPRAPGDVLHDAPPAS 110
DB 52 --NKIYSGDITTLRUFINRYTRDHLRVRHITDAKQORWEVPYNLLRREPPNVI---GKSR 106

```



**RESULT 8**  
**A32609**  
 alpha-glucosidase (EC 3.2.1.20) precursor, lysosomal - human  
 N;Alternate names: acid alpha-glucosidase; acid maltase; maltase-glucoamylase  
 C;Species: Homo sapiens (man)  
 C;Date: 21-May-1990 #sequence\_revision 05-Apr-1995 #text\_change 18-Jun-1999  
 C;Accession: A40577; A32609; S00831; S18847; I52309; S63526  
 R;Martiniuk, F.; Bodkin, M.; Tzall, S.; Hirschhorn, R.  
 DNA Cell Biol. 10, 283-292, 1991  
 A;Title: Isolation and partial characterization of the structural gene for human acid alpha-glucosidase  
 A;Reference number: A40577; MUID:91229698; PMID:1674202  
 A;Accession: A40577  
 A;Status: not compared with conceptual translation  
 A;Molecule type: DNA  
 A;Residues: 1-952 <MAR>  
 A;Note: order of exons in Figure 2 is incorrect  
 R;Martiniuk, F.; Mehler, M.; Tzall, S.; Meredith, G.; Hirschhorn, R.  
 DNA Cell Biol. 9, 85-94, 1990  
 A;Title: Sequence of the cDNA and 5'-flanking region for human acid alpha-glucosidase, cDNA with previous cDNA and amino acid sequences.  
 A;Reference number: A32609; MUID:90262651; PMID:2111708  
 A;Accession: A32609  
 A;Molecule type: mRNA  
 A;Residues: 1-952 <MA2>  
 A;Cross-references: GB:M34424; NID:g182907; PIDN:AAAS2506.1; PID:g182908  
 R;Martiniuk, F.; Bodkin, M.; Tzall, S.; Hirschhorn, R.  
 Am. J. Hum. Genet. 47, 440-445, 1990  
 A;Title: Identification of the base-pair substitution responsible for a human acid alpha-glucosidase  
 A;Reference number: A35698; MUID:90365036; PMID:2203258  
 A;Contents: partial sequence of GAA 2 allelic form  
 A;Accession: A35698  
 A;Status: not compared with conceptual translation  
 A;Molecule type: mRNA  
 A;Residues: 1-90, 'N', 92-129 <MA3>  
 R;Hoefloot, L.H.; Hoogveen-Westerveld, M.; Kroos, M.A.; van Beeumen, J.; Reuser, A.J.J.  
 EMBO J. 7, 1697-1704, 1988  
 A;Title: Primary structure and processing of lysosomal alpha-glucosidase; homology with  
 A;Reference number: S00831; MUID:89005058; PMID:3049072  
 A;Accession: S00831  
 A;Molecule type: mRNA  
 A;Residues: 1-126, 'VLL', 130-198, 'R', 200-222, 'H', 224-371, 'L', 373-401, 'R', 403-412, 'TSRSTRM'  
 A;Cross-references: EMBL:Y00839  
 A;Accession: S18847  
 A;Molecule type: protein  
 A;Residues: 70-89, 123-126, 'VLL', 130-145, 204-215, 230-249, 332-345, 349-370, 394-401, 'R', 403-412, 'TSRSTRM'  
 R;Lin, C.Y.; Shieh, J.J.  
 Biochem. Biophys. Res. Commun. 208, 886-893, 1995  
 A;Title: Identification of a de novo point mutation resulting in infantile form of Pompe  
 A;Reference number: I52309; MUID:95209708; PMID:7695647  
 A;Accession: I52309  
 A;Status: preliminary; translated from GB/ENBL/DBJ  
 A;Molecule type: mRNA  
 A;Residues: 631-644, 'H', 646-680 <RES>  
 A;Cross-references: GB:S76893; NID:g912786; PIDN:AAB33842.1; PID:g912787  
 A;Note: homozygous mutation of Asp-645 to His causes an infantile form of Pompe's disease  
 R;Fuller, M.; van der Ploeg, A.; Reuser, A.J.J.; Anson, D.S.; Hopwood, J.J.  
 Eur. J. Biochem. 234, 903-909, 1995  
 A;Title: Isolation and characterization of a recombinant, precursor form of lysosomal acid  
 A;Reference number: S63526; MUID:96163476; PMID:8575451  
 A;Accession: S63526  
 A;Molecule type: protein  
 A;Residues: 68-74, 'X', 76-79, 'X', 81 <FUL>  
 C;Comment: In common with other lysosomal enzymes, posttranslational processing includes processing at both amino and carboxyl ends.  
 C;Genetics:  
 A;Gene: GDB:GAA  
 A;Cross-references: GDB:119965; OMIM:232300  
 A;Map position: 17q23-17q23  
 A;Introns: 182/3; 231/2; 286/3; 319/1; 359/1; 398/3; 442/3; 479/3; 517/3; 546/1; 595/2;  
 A;Note: GAA 1, GAA 2, and GAA 4 are common alleles in the normal population

Dd	868	LERGAYTQVIFLARNTIVNE--LVRVTSE-----GAGLQKQKTVLGG--VATAPOQ 915	
Qy	827	LTVHNSAEYASS---SAGTRYONAGGLGVGVAHIGLSLVWGEFF 869	
Dd	916	---VLSNGVPVSNFTYSPDKVLDI-----CVSLLMGEOQ 947	
RESULT 9			
A23945	sucrose alpha-glucosidase (EC 3.2.1.48) / oligo-1,6-glucosidase (EC 3.2.1.10) - rabbit		
N:	Alternate names: small intestinal sucrase/isomaltase (SI)		
C:	Species: Oryctolagus cuniculus (domestic rabbit)		
C:	Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 24-May-1996		
C:	Accession: A23945; B25987; A29163		
R:	Hunziker, W.; Spiess, M.; Semenza, G.; Lodish, H.F.		
Cell	46, 227-234, 1986		
A:	Title: The sucrase-isomaltase complex: primary structure, membrane-orientation, and ev		
A:	Reference number: A23945; MUID:86245068; PMID:3755079		
A:	Accession: A23945		
A:	Molecule type: mRNA		
A:	Residues: 1-1827 <HUN>		
R:	Sjoestrom, H.; Noren, O.; Christiansen, L.A.; Wacker, H.; Spiess, M.; Bigler-Meier, B		
FEBS	Lett. 148, 321-325, 1982		
A:	Title: N-terminal sequences of pig intestinal sucrase-isomaltase and pro-sucrase-isoma		
A:	Reference number: A25987; MUID:83105704; PMID:7152027		
A:	Accession: B25987		
A:	Molecule type: protein		
A:	Residues: 2-32, 'XXX', 36-38, 1008, 'N', 1010-1014, 'E' <SJ2>		
R:	Frank, G.; Brunner, J.; Jauser, H.; Wacker, H.; Semenza, G.; Zuber, H.		
FEBS	Lett. 96, 183-188, 1978		
A:	Title: The hydrophobic anchor of small-intestinal sucrase-isomaltase. N-terminal sequ		
A:	Reference number: A29163; MUID:79086207; PMID:729784		
A:	Accession: A29163		
A:	Molecule type: protein		
A:	Residues: 2, 'VNA', 6-32, 'XXX', 36-38 <FRA>		
C:	Comment: Carbohydrate analysis of the mature enzyme complex indicates both N- and O-li		
C:	Complex: the two product chains remain associated after cleavage		
C:	Function: <ISM>		
A:	Description: oligo-1,6-glucosidase catalyzes the hydrolysis of 1,6-alpha-D-glucosidic		
A:	Pathway: carbohydrate digestion		
C:	Function: <SUC>		
A:	Description: sucrose alpha-glucosidase catalyzes the hydrolysis of the 1,2-alpha-D-glu		
A:	Pathway: carbohydrate digestion		
C:	Superfamily: sucrase/isomaltase; sucrase/isomaltase homology; trefoil homology		
C:	Keywords: carbohydrate digestion; duplication; glycoprotein; glycosidase; hydrolase; m		
F:	2-1007/Product: oligo-1,6-glucosidase (isomaltase chain) #status experimental <ISO>		
F:	13-32/Domain: membrane associated #status predicted <TM>		
F:	43-65/Region: serine/threonine-rich		
F:	63-109/Domain: trefoil homology <TRF1>		
F:	189-840/Domain: sucrase/isomaltase homology <SIM>		
F:	931-977/Domain: trefoil homology <TRF2>		
F:	1008-1827/Product: sucrose alpha-glucosidase (sucrose chain) #status experimental <SUC		
F:	1062-1734/Domain: sucrase/isomaltase homology <SIM2>		
F:	12/Binding site: carbohydrate (Thr) (covalent) #status experimental		
F:	99,455,859,896,904,1235,1303,1325,1340,1354,1368,1403,1535,1572,1748,1763,1799/Binding		
F:	1007-1008/Cleavage site: Arg-11e (trypsin) #status predicted		
Query Match 29.6%; Score 1370; DB 1; Length 1827;			
Best Local Similarity 36.9%; Pred. No. 2e-95;			
Matches 316; Conservative 142; Mismatches 303; Indels 96; Gaps 25;			
Qy	54	RSST-----GDVQRLAVVASLETDSRLVRITDADHPWEVPQDIIPR---PAPGDVLHD 105	
Dd	132	RKSTPTLFGNDINNVLTTESQTANRLRFLKLTDPNNKRYEYFPHQVTFEPFAGPAATETLYD 191	
Qy	106	APPASAPLOGRVLSPPAGSDLVLTVAHSPFRFTVSRSTGTLTDTAPG-LVFPDKYLEV 164	
Dd	192	-----VQVTENPFSSIKVIRKNNRILFDSSIGLVYSDQLQI 229	
Qy	165	TSALPAGRASLYGLGHETKSFLRHRNDSFTLWADIGASYVD---VNLVYSGHPFYMDVR 221	
Dd	230	STRFLPS--EYMYGFEHVKH--RFRHDLVYKWTPIFTTRDQHTDNNNNLYGHQTFFNCIE 285	

QY	222	-APCTAHGVLLSSNGDV-LYGSVYTKYKVGVLDFEYFAGBNPLAVVDQYQTLIARP 279	
Dd	286	DTTGKSGFVFLMNSNAMEIFIQPTPIVTVYVIGILDYFIPLGDTPEQVVQVOELGRP 345	
QY	280	APMEYWSFGHQCRYGLVNSDLERVVARYAKARIPLEVMMTIDIDYMDGFKDFTLDRVNF 339	
Dd	346	AMPAYMSLGFLSRWYNLSLDVVKVVRNREALIPDTCVSDIDYMEDKKDFTYDRVAY 405	
QY	340	TAEELRPFDLHRNAQYVILDPGRVD--PIDATYGTFFVGMQODIFLKNNG--TNF 395	
Dd	406	NG--LPDFVQDLHGQKQYIILDPALISINRRASGEAYESYDRGNAQNVWVNESDGTPI 463	
QY	396	VGNVMPGDVYPPDFHMAAEFAREISLFRRTIPVDGLMIDMNEISNFYNPEPM---N 451	
Dd	464	VGEVMPGDTVPDFTSPNCIEWANECEIHOEVYVYDGLMIDMNEVSSFFVQSGNKGNDN 523	
QY	452	ALDDPPYKINNDDGTRPINNKTVPLAVHYGGVTYEENHFLGLEAPATGRGVLR--DTG 510	
Dd	524	TLNTPPY--IPDIVDKLMSKTLQMSQVYWG-KQYDVHSLYGYSAIATERAVERVFPN 580	
QY	511	RPFVLSRSTFVSGRYTAYWTGDNAAATWGLRLYSINTMLSFGFLGMPMIGADICGFNGN 570	
Dd	581	KRSFILTRSTFAGSGRAAAHWLGONTATWQEMWSITGMLEFGLFGMPLVGADICGQAE 640	
QY	571	TTEELCGRWIOLGAFYFSPDRHSA-----IFTVRRELYLWPSVAASGRKALGLRYQL 622	
Dd	641	TTEELCRWMLGAFYFSPRNHNDGFEHQDPAFFGQDSL-----LVKSSRHYLNIRYTL 695	
QY	623	LPYFTLMEAHMTGAPLARPLFFSYPHDVATYVDQFLGLRGVLSVPLEPQFTTVA 682	
Dd	696	LPPLYTLFYKAHGETVAREVLEHEFYEDTNSWYEDREFLWGPALLITPVLTQGAETVSA 755	
QY	683	YFPAGRWVRLYDLSLAVATRTG-----KHVRUPAPADTVVHVTGTTLPLOQSALT 735	
Dd	756	YIPDAVW--YDY-----ETGAKRPRKQRENSLPADKLGHLRGYIIPQPAVTT 806	
QY	736	SRARTAPHLVALAEDGTASGYILFDGDSPEYRRSDMSWVFNFKYIPNNKGAIKVKS 795	
Dd	807	TASRNNPLGLIILALDDNDTAVGDFPDDGETKDTQVNDNY--ILYTFVSNM---NLNI 860	
QY	796	EVVHNSYAQSTLIVSKVLVGMHSPAPKGLTVHNSAEVEASSAGTRVQNAAGLGGV 855	
Dd	861	TCHELISEGTTFLAQTIKILGVTTVT--QVTVAENQSMSTHNTFYDPSNQ-----V 913	
QY	856	AHIGGLSLVWGEFEELK 872	
Dd	914	LLIENLNFNLGRNFRVQ 930	
RESULT 10			
JC4217			
alpha-glucosidase (EC 3.2.1.20) - Aspergillus oryzae			
C:Species: Aspergillus oryzae			
C:Date: 21-Nov-1995 #sequence_revision 08-Feb-1996 #text_change 21-Jul-2000			
C:Accession: JC4217			
R:Minetoki, T.; Gomi, K.; Kitamoto, K.; Kumagai, C.; Tamura, G.			
BioSci. Biotechnol. Biochem. 59, 1516-1521, 1995			
A:Title: Nucleotide sequence and expression of alpha-glucosidase-encoding gene (agda)			
A:Reference number: JC4217; MUID:96032211; PMID:7549103			
A:Accession: JC4217			
A:Molecule type: DNA			
A:Residues: 1-985 <MIN>			
A:Cross-references: DBBJ:D45179; NID:g1054564; PIDN:BAA08125.1; PID:g1054565			
C:Comment: This enzyme catalyzes the liberation of alpha-glucose from the non-reducin			
portant enzyme in the food industry as the isomalto-oligosaccharides.			
C:Genetics:			
A:Gene: agda			
A:Introns: 234/2; 371/2; 428/2			
C:Superfamily: Schwannomyces glucan 1,4-alpha-glucosidase GAM1; sucrase/isomaltase h			
C:Keywords: glycoprotein; glycosidase; hydrolase			
F:152-898/Domain: sucrase/isomaltase homology <SIM>			
F:126.145,255,349,424,508,536,539,602,624,661,835,881,929,957/Binding site: carbohydr			
F:492/Active site: Asp #status predicted			



Query Match	29.1%	Score 1348.5	DB 2	Length 985
Best Local Similarity	34.1%	Pred. No. 3.4e-94		
Matches	312	Conservative 158	Mismatches 289	Indels 157
Gaps	28			
QY	59	DVORLAVASLETDSRLVRIT----	DADHPRWEV-PODIIPRPAQDVLHDPASSAP	113
Db	97	DVDSLTTLTVEYQAKRLNIQIVPTFDASNAWYLSSELPFRP-----	KASQNASVP	149
QY	114	LQGRVLSPAGSDVLTVHASP--FRETVSRRSTGDTLPT-APGLVFRDKYLEVTSALPAG	171	
Db	150	-----QSDPVVWSNEPSFNKVKATGDVLFNTKSTLYVENQPIEFVLLPE-	199	
QY	172	RASLYGLGHTKSSFLRHNSFTLWADIGASYVDVNLVYGHSHPFYMDVR-----	221	
Db	200	EYNLYGLGER-NKQLRLLENANLTYAADI-ADPIDNNIYGHAFYLDTRYKYVGQONKS	257	
QY	222	-----APCT-----AHGVLLSSNGMVLVYGSVYVYKVGVLDFEYFAGNPPLA	267	
Db	258	HTVKSSEAPSQEVSYSHGVFLRNAGQELLRDQKLIWRTLGSSVDLTFYSGPTQAE	317	
QY	268	VYDQY-TQLIARPAMPYMSFGHCQRYGLNVSDLERVARYAKARIPLEVMTDIDYM	326	
Db	318	VTKYQVLSVGLPAMQYNTLGFHCQRYGNWSEFEDVLNFERFPIPLVWADIDYM	377	
QY	327	DGFKDFTLDNRVNTAAELRPFVDRLHNAQKYLILDPGIRV-DPIDAT--YGTFTVRGMQ	383	
Db	378	HGYNFNDNDQHRFSYEGEKKLNLKUHAGGRWVP-IVDGALYIPNENASDAYETVDRGAK	437	
QY	384	QDIFLAK-NGTNFVGNVPGDVFYDFMHPAAEFWAREISLFRRTIPVDGLMTDMNEIS	442	
Db	438	DVDFIKNPGSLYIGAVMFGYTVYDWHHPKASDFWANELVTWNKLNHYDGVWYDMAEVS	497	
QY	443	NF-----YNP-----EPNNA-----	452	
Db	498	SFCVSGCGTGNLSMNPAPFPFALPQEPGNVYVYDPEGFNITNATEAASASAGAAQSAAA	557	
QY	453	-----LDDPPVIRINDGTGRPINNKTVPLAVHYGVVTEYEEHNL	492	
Db	558	SSTTSAPYLRTPTPGVNRVDHPPVYVNHVQVGHDLVSHVISPSTHSDGVQYDVHSL	617	
QY	493	FGLLEARATGRGLRD-TGRREFVLSRSTFVSGRYTAYWTGDNAATWGLDLYRSINTMLS	551	
Db	618	YCHOGINATYHGLLKWENKRFPIIARSTFSGSKWAGHWGDNFSGWMSFFSISOALQ	677	
QY	552	FGLFGPMNIGALICGPNWNTTELGRWLOLQAFYPSFSDHSAITVRELYLPSVAAS	611	
Db	678	FSLFGIPMFQVDTGCGFNGNTDELCNRWQLSAFFFPFYNNHNVLSAIPQEPYRWASVIDA	737	
QY	612	GRKALGRYQLLPYFYTLMYEAHMTCAPIARLPFFSYPRHDVATYGVDRQFLLGRGLVSP	671	
Db	738	TKAAMNIRVAILPYFYTLFHLAHTGSTVWRBALAWEFNDPDSLAAGVQTLVGPSSVMVIP	797	
QY	672	VLEPGPTVDVAPPA---GR-WYRLYDYSLAVATRTGKVRLPAPADVTWNHLTGTLIP	727	
Db	798	VLEPQVDTVQGVPPGVGHEGWYDWSQT-AVDAPKPGVNTTTSAPLGHIPFVRCGSLP	856	
QY	728	LOQSALTTSRRARTAFHLLVALAEDCTAGXYLFLDDGDS--PEYGRSDWSMVRENYKIP	785	
Db	857	MQEVALTTRDARKTPSLASLSSNGTAGSGLYLDGGSVYFEDLSVDLFLASRSLR-A	915	
QY	786	NNKGAIKVKSSEVHNSYAQSRTLVIKSVLMGHRSPAPKLTLVHNSAEVSAESSAGTR	845	
Db	916	SARGTWKEANP-----LANVTVLG-----VTEKPSVSVTING---ETLSDSDSVK	955	
QY	846	YQNAGGLGGVAHIGGL	861	
Db	956	Y---NATSHVLHVGGGL	968	

RESULT 11

UUHU

sucrose alpha-glucosidase (EC 3.2.1.48) / oligo-1,6-glucosidase (EC 3.2.1.10) [validated]

N:Alternate names: limit dextrinase; small intestinal sucrase/isomaltase (SI)  
C:Species: Homo sapiens (man)  
C:Date: 19-Nov-1988 #sequence revision 24-May-1996 #text\_change 08-Dec-2000  
C:Accession: S36082; A27326; S24329; A61136  
R:Lacasa, M.  
submitted to the EMBL Data Library, December 1991  
A:Reference number: S36082  
A:Accession: S36082  
A:Molecule type: mRNA  
A:Residues: 1-1827 <LAC>  
A:Cross-references: EMBL:X63597; NID:G36644; PIDN:CAA45140.1; PID:G36645  
R:Green, F.; Edwards, Y.; Hauri, H.P.; Povey, S.; Ho, M.W.; Pinto, M.; Swallow, D.  
Gene 57, 101-110, 1987  
A:Title: Isolation of a cDNA probe for a human jejunal brush-border hydrolase, sucras  
A:Reference number: A27326; MUID:88112852; PMID:2962903  
A:Accession: A27326  
A:Molecule type: mRNA  
A:Residues: 1-661, 'X', 663-678 <GRE>  
A:Cross-references: GB:M22616  
R:Chantret, I.; Lacasa, M.; Chevalier, G.; Ruf, J.; Islam, I.; Mantei, N.; Edwards, Y.  
Biochem. J. 285, 915-923, 1992  
A:Title: Sequence of the complete cDNA and the 5' structure of the human sucrase-isom  
A:Reference number: S24329; MUID:92359963; PMID:1353958  
A:Accession: S24329  
A:Molecule type: mRNA  
A:Residues: 1-661, 'P', 663-931 <CHA>  
A:Cross-references: EMBL:X63597  
R:Gorvel, J.P.; Ferrero, A.; Chambrault, L.; Rigal, A.; Bonicel, J.; Maroux, S.  
Gastroenterology 101, 618-625, 1991  
A:Title: Expression of sucrase-isomaltase and dipeptidylpeptidase IV in human small i  
A:Reference number: A61136; MUID:91317403; PMID:1677636  
A:Accession: A61136  
A:Molecule type: protein  
A:Residues: 2-14, 'P', 16-20; 1008-1015, 'E', 1017-1021, 'TX', 1024 <GOR>  
C:Genetics:  
A:Gene: GDB:SI  
A:Cross-references: GDB:120377; OMIM:222900  
A:Map position: 3q25.2-3q26.2  
C:Complex: the two product chains remain associated after cleavage  
C:Function: <ISM>  
A:Description: oligo-1,6-glucosidase catalyzes the hydrolysis of 1,6-alpha-D-glucosid  
A:Pathway: carbohydrate digestion  
C:Function: <SUC>  
A:Description: sucrose alpha-glucosidase catalyzes the hydrolysis of the 1,2-alpha-D-  
A:Pathway: carbohydrate digestion  
C:Superfamily: sucrase/isomaltase; sucrase/isomaltase homology; trefoil homology  
C:Keywords: carbohydrate digestion; duplication; glycoprotein; glycosidase; hydrolase  
F:2-1007/Product: oligo-1,6-glucosidase (isomaltase chain) #status experimental <MATI  
F:13-32/Domains: membrane associated #status predicted <TMM>  
F:42-60/Region: serine/threonine-rich  
F:63-109/Domains: trefoil homology <TRF1>  
F:189-840/Domains: sucrase/isomaltase homology <SIM>  
F:931-977/Domains: trefoil homology <TRF2>  
F:1008-1827/Product: sucrose alpha-glucosidase (sucrase chain) #status experimental <  
F:1062-1734/Domains: sucrase/isomaltase homology <SIM2>  
F:12/Binding site: carbohydrate (Ser) (covalent) #status absent  
F:99,437,455,823,855,904,926,1235,1303,1340,1354,1403,1535,1572,1675,1748,1763,1815/B  
F:1007-1008/Cleavage site: Arg-11e (trypsin) #status predicted

Query Match 28.7% Score 1329.5; DB 1; Length 1827;  
Best Local Similarity 36.4%; Pred. No. 2.4e-92;  
Matches 307; Conservative 144; Mismatches 309; Indels 83; Gaps 23;

QY	59	DVORLAVASLETDSRLVRITDADHPRWEVPODIIPR----	PAPGDVLHDPASSAPLQ	115
Db	142	DINSVLFTTQNTQTPNFRFKITDPNNRVEVPHQVKEFTGTVSDTLVD-----	191	
QY	116	GRVLSPAGSDVLTVHASPFRFTVRRSTGDTLFTAPG-LVFRDKYLEVTSALPAGAS	174	
Db	192	-----VKVAQNPFSIQVIRKSNKGLTFTSIGFLVYSQYLQISARLPSD--Y	237	
QY	175	LYGLGHTKSSFLRHNSFTLW-----NADIGASYVDVNLVYGHSHPFYMDVR-APGTAGVUL	230	

Db 238 IYIGEQVHK--RFRHLSWKTWPIPTRDQLQPDNNNNLYCHQTFMCIEDTSGKSGVF 295  
QY 231 LLSNGMDV--LYGGSYVYKVGGLDFFYFAGPNPLAVVDQYTQLIARAPMYSFGF 289  
Db 296 LMSNAMEIFIQTPIVYRTVGILDFYLLGDTPEQVQQYQQLVGLUPAMPAYWNLGF 355  
QY 290 HOCRYGVLNDSLERVAVYAKARIPLEVNMWTDIDYMDGFKDFTLDRVNFATAELPFFVD 349  
Db 356 QLSRWNYKSLDVVKEVVRNREAGIPDQTVTIDIDYMDGFKDFTLDRVNFATAELPFFVD 413  
QY 350 RLHNAQKYLILDPGRVD--PIDATYGTFRGMQODIFLKN--GTNPGVNWFGDYY 405  
Db 414 DLHDGQYVIIIDPAISICRRANGTYATYERGNTQHWINESDGSPTPIGFWPGLTV 473  
QY 406 FPDFMHPAAEFWAREISLFRRTIPVDGLWIDMNEISNEYNP--MNALDDPPYRIN 461  
Db 474 YPFTNPNCIDWANECSIFHOEVOYDGLWIDMNEVSSFQGTGKGNVKNLNPFF--T 531  
QY 462 NDGTGRPINNKTVPLAVHYGGVTEYBEHNLFLGLLEARATGRGLR--DTGRRPFFVLSRST 520  
Db 532 PDILDKLMSKTIAMDVQNG--KQYDVHSLYGYMAIATEQAQVQKVPNKRFSILTRST 590  
QY 521 FVSGRYTATWTGNAATGDLKYSINTMLSFGLFGMPMIGADICGPNNGTTEELCGRWI 580  
Db 591 FAGSGRHAHMLGDNATASMEQMEWSITGMLFSLFGIPLVGADICGFVAETTEELCRWM 650  
QY 581 QLCAFYFESRDHSA-----IFTVRRELYLWPSVAASGRKALGRYOLLPYFTLMYE 632  
Db 651 QLCAFYFESRNHNSDGYEHODPAFFGNSL-----LVKSRQYLTIRYLLPPLYTLFYK 705  
QY 633 AHMTGAPIARPLFPFSPHVDATYGVDRQFLGLGRVLVSPVLEPGTTPVDAYFPAGRWRL 692  
Db 706 AHVFGETVARPLHEFYEDTNSWIEDTEFLWGPALLITPVLKQADTVSAIYDAW--- 762  
QY 693 YDY-SLAVATRTGKHVRLPAPADTVNHLTGGTILPQSQALTTSRARATFAHLLVALAE 751  
Db 763 YDYESAKRPMRKQVDMYLPADKIGLHGRGGYIIPQEPDVTITASRKNPLGLIVALGE 822  
QY 752 DGTASGVFLDDGDSPEYGRSDMSVMVRFNYKIPNKGAIKVKSEVVHNSYAQSRTLVIS 811  
Db 823 NNTAKGDFWDDGETKDTQNGY--ILYTFVSNN-----TLDIVCTHSSYQGTTLAQ 876  
QY 812 KVLGMGRSPAAPKLTLVHNSAEVASSAGTRYQN--AGLGGVVAHIGGLSLVVGEEF 869  
Db 877 TVKILG-----LTDSTVEVRAENQPMNAHSNFTYDASNOVLLIADUKLNGRF 927  
QY 870 ELX 872  
Db 928 SVQ 930

RESULT 12  
T50267  
probable family 31 glucosidase [imported] - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 21-Jul-2000  
C:Accession: T50267  
R:Hunt, C.; Aves, S.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, December 1999  
A:Reference number: Z25031  
A:Accession: T50267  
A:Status: preliminary;  
A:Molecule type: DNA  
A:Residues: 1-995 <HUN>  
A:Cross-references: EMBL:AL133522; PIDN:CAB63549.1; GSPDB:GN00066; SPDB:SPAC922.02c  
A:Experimental source: strain 972h(-); cosmid c922  
C:Genetics:  
A:Gene: SPAC1039.1lc; SPDB:SPAC922.02c  
A:Map position: 1  
C:Superfamily: Schwanniomyces glucan 1,4-alpha-glucosidase GAMI; sucrose/isomaltase hom

Query Match 28.1%; Score 1302; DB 2; Length 995;  
Best Local Similarity 34.0%; Pred. No. 1.2e-90;

Matches 304; Conservative 153; Mismatches 285; Indels 152; Gaps 26;  
QY 59 DVQRLAVVASLETDSRLRVRTDADHPWEVPQDIIPRAPCDVLDAPPASSAPLOG-R 117  
Db 108 DYPFLALANVSFSDIRLHVSIQDLYGAQOFQSKRT-----DV-WDAPLYHFQFQGD 159  
QY 118 VLSPAGSDLVLTVAHPPFRTVSRSTGDTLFDOT--APGLVFRDKYLEVTLSALPAGRASLY 176  
Db 160 TYN-----FSNPQPFVFWTRVSDGEVLFDTRGHKLIFEDQYIELTNN--VDYVNY 211  
QY 177 GLGETKSFRLRHNDSTFLNADIGASYVDNLYGSHPPFYMDVR-----A 222  
Db 212 GLAE--TVHGLRLGNLTFTF--ANGNPTPLDRNAVGTHTFFYLEHRYTPSENLSNGQPSY 269  
QY 223 PCTAHGVLILLSNGMDVLYGGSYTYKYVIGVLDYFFAG--PNPLAVVDQYTQLIARPA 280  
Db 270 TSSTHGVLMLTANGKEVLLRPNYLQYRIIGIVDLYIYVGGTKNPKDKTVSFQVSGVTPA 329  
QY 281 PNPYMSFGHOCRYGLVANSDLERVVARYAKARIPLEVNMWTDIDYMDGFKDFTLDRVNF 340  
Db 330 MQOHWTFGFHICRWGYNVFDLVEKENPKFEPDVTFNDSIDIDMYEYRFTVESNAPP 389  
QY 341 AAELEPFFVDRLHRNAQKYLILDPGI--RVDPi---DATYGTFRGMQODIFLKR--NGTNF 395  
Db 390 KDKMFEFNSLQOSNOHYPIIDAAIYAANPINRSDDVVYVYEGVRRDIFLRFNDRSLY 449  
QY 396 VGNVWPGDVYFPDMHPAAAEFWAREI-----SLFRRTIPVDGLWIDMNE----- 440  
Db 450 VGNVWPGFTTTPDFTNPTTYTECLMNLSSAAGYNSSF--PLPYSLGWLIDMNEPTSF 507  
QY 441 -----ISNFI----- 445  
Db 508 IGSCGTDKLDQNPVHPAFILGESEPNMVMYVPEGFEHTNASEHASAYQASVYATATS 567  
QY 446 -----NPEPMNALDDPPYRINNDGTGRPINNKTVRPLAVHYGGVTEYEEHNL 493  
Db 568 TVESVKATSTPLNVRPKYNINVPYALNTEQEGDLSNLGVSVNATYHDTGVRYNLFNTY 627  
QY 494 GLLEBARATGRGLR--DTGRRPPLVSRSTFVSGRYTATWTGDNAAATGDLKYSINTMLSP 552  
Db 628 GYDQSRVYDSTLTSTIEPNVRPFIILSRSTFVSGRYTAAHMLGDNYSLSWNNMIFSPGALT 687  
QY 553 GLFGMPMIGADICGPNNGTTEELCGRWIOLGAFYFESRDHSAIPTVRRELYLWPSVAASG 612  
Db 688 NMVGLPMVADVCGPNGMTDEELCSRMMALAFIPFYHNSLGSISQEPFRWESVASS 747  
QY 613 RKALGLRYOLLPYFTLMYEAHMTGAPIARPLFPSPHVDATYGVDRQFLGLGRGVLSVP 672  
Db 748 RCAMNIRYSLLPYWTLMYEAASSQGLPLIRPLFEFPNEPSLANADRFVWGSALLVTPV 807  
QY 673 LEPGTTVDAYFPAGR---WYRLYDYSLAVATRTGKHVRLPAPADTVNHLTGGTILPQ 729  
Db 808 LEPNVYVRGVPFGDNSTWYMDYDHKV--IYRQHNENITLSAPLTHINVAIRGGNIIPMQ 866  
QY 730 QSALTTSRRARTAFHLLVALAEDGTASGYFLDDGDSPEYGRSDMSVMVRFNYKIPNKG 789  
Db 867 KPSLTHTHTKQNPDLVALDSDRACGSLYVDDGVSQI---QESTLFXF-----VANGD 919  
QY 790 AIKVKSE---VHNSYAQSRTLVIKVVLMGRSPAAPKLTIVHNSAEVASS 840  
Db 920 SLSIESYGDQLQWHEP-----LSKITIIG--LPCAP--IGVYFEGVQVESFS 961

RESULT 13  
T38598  
probable family 31 glycosyl hydrolase (alpha glucosidase) precursor - fission yeast  
C:Species: Schizosaccharomyces pombe  
C:Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 20-Oct-2000  
C:Accession: T38598; S62559; T38911  
R:Pearson, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.  
submitted to the EMBL Data Library, November 1995  
A:Reference number: Z21801  
A:Accession: T38598

A:Molecule type: DNA  
A:Residues: 1-385 <PEA>  
A:Cross-references: EMBL:Z67961; PIDN:CAA91887.1; GSPDB:GN00066; SPDB:SPAC30D11.01c  
A:Experimental source: strain 972h-; cosmid c30D11  
R:Pearson, D.; Churcher, C.M.  
submitted to the EMBL Data Library, November 1995  
A:Reference number: S62559  
A:Accession: S62559  
A:Molecule type: DNA  
A:Residues: 1-384 <PE2>  
A:Cross-references: EMBL:Z67961  
R:Pearson, D.; Churcher, C.M.; Bartelli, B.G.; Rajandream, M.A.; Walsh, S.V.  
submitted to the EMBL Data Library, February 1996  
A:Reference number: Z21817  
A:Accession: T38911  
A:Molecule type: DNA  
A:Residues: 352-993 <PE3>  
A:Cross-references: EMBL:Z69728; PIDN:CAA93572.1; GSPDB:GN00066; SPDB:SPAC56F8.0.  
A:Experimental source: strain 972h-; cosmid c56F8  
C:Genetics:  
A:Gene: SPDB:SPAC30D11.01c; SPDB:SPAC56F8.01  
A:Map position: 11  
C:Superfamily: Schwanniomyces glucan 1,4-alpha-glucosidase GAM1; sucrose/isomaltase homology (fragment) <SIM>  
F:165-384/Domain: sucrose/isomaltase homology (fragment) <SIM>

Query Match 27.7%; Score 1282.5; DB 2; Length 993;  
Best Local Similarity 32.9%; Pred. No. 3.7e-89;  
Matches 286; Conservative 156; Mismatches 287; Indels 141; Gaps 20;

QY 59 DVORLAVASLETSLRLVRITADHPHWRVQDIIIPRAPGDVLDHAPPASSAPLOGRV 118  
DB 114 DYPVLLNVSYDTEERHVISIDLNQTFQ-----LSNRDWDADPLFYRS 159

QY 119 LSPAGS-DLVLTVHASPRFTVSRSTGDTLFDTPAG-LVFRDKYLEVTSALPAGRASLY 176  
DB 160 SNFGNLOYNFSNTDPEFWITRIADQVLFDRGNPLIFEDQVIELTNM-VEDYNYV 218

QY 177 GLGEHTKSSFLRHNDSTFLNADIGASYDVNLVGHGPPFMDVR-----APGTAH 227  
DB 219 GL-SGSQSFRLGNLTLTFWATGYSDS-PEANMYGSHPFMEQRYPIGTNTYTSASH 276

QY 228 GVLLSSNGMDVLYGGSVYTKVIGVLDVFFPAGP--NPLAVVDQYTLIARPAHPYW 285  
DB 277 GVLLSSNGMEVLLRSTYIKYRMIGIIDLFVYSGTSPKTYIQVYQSIGTPTMQPYW 336

QY 286 SFGHQCRGYGLNVSDLRVAVAKARIPLEVMTDIDVMDGKDFLDRVNFATAELR 345  
DB 337 SLGFQMSRWGYKTLSDLINMRSYLNASNIPTGFWNDIDYSEFRTFYVNSTAFPPNQT 396

QY 346 PFVDRLRHNAQYVLIIDPGI-RVDP-----IDATYGFVRGMQDIFLKR-NGTNFVGNVW 400  
DB 397 DFFRSLDESHQHYVPVLDPAIYAANPKSADRTYYPYSGFEDNIFIKNPNGSAYVGMW 456

QY 401 PGDVFPDFMHPAAAEFAWEI-----SLFRTTIPDGLWIDMNEISNF----- 444  
DB 457 PGFVYVPDFTNPVLYQWKGQILNLSAFGNYSYDLFPFSLGLCLDMNEPTSCFICGSD 516

QY 445 -----YN-PEPMNA----- 452  
DB 517 LLKLNVPVPPFSLPGDVNKNVSYPEDFNATNTTEYKSVRASOSQYKATATSEKSHETP 576

QY 453 -----LDOPPYRINNDGTGRPINNKTVRPLAVHYGGVTEYEEHNLFGLLLEAR 499  
DB 577 SSESILINGKPEFSINYPVALDQTEHDLAQPGVSPNATWGHNTLRVNLFTNYGSESK 636

QY 500 ATGRGVLR-DTGRRPVLSRSTFGSGRYTAYWTDNAAATGDLRYISINTMLSGLFQMP 558  
DB 637 ISFEALNSTQPNIRPFLLSRSTFGSGRYAAHMLGDNKSQWSDMVSSTSSILTENLLGIP 696

QY 559 MTGADICGNGNNTTEELGRWTLQLCAFYFSPRDHGAIFTVRRELYLWPSVAASGRKALGL 618  
DB 697 MVGADVCGINGNTDELCAARMALGALFFPYRNHNSLGSIFOEPRFASVAEASRSIAEI 756

QY 619 RYQLLPYPTLMEYEAHMTGAPIARPLFPSPHYDVATYGVDRQFLLRGVLYSPVLEPGPT 678  
DB 757 RYSLPLPYPTLMEYEAHMTGAPIARPLFPSPHYDVATYGVDRQFLLRGVLYSPVLEPGPT 816

QY 679 TVDAYFPAGR---WYRLDYSLAVATRTGKVRLPAPADTVNVHLTGCTILPQOQSALT 735  
DB 817 YIQGIIPGDNQDITWYDYNHVS-INHDYDENITWSAPLGYNIAVRGNNIIPQOPGYTT 875

QY 736 SPARTAFHLLVALAEDGTASGYLFDGDSPEYGRKSDMSVRFNPKIPNNKGAIKVKS 795  
DB 876 YESRNNPYSLLITAMNNGFASGSLYIDDGISNQ-----TNSSLVSKLNS 919

QY 796 EVVHNSYAQSRLV-----ISKVLMGHESP 821  
DB 920 NSNTITCVVGTWVSSPSLANITILGLSNP 949

RESULT 14  
JN0102  
Glucan 1,4-alpha-glucosidase (EC 3.2.1.3) GAM1 precursor - yeast (Schwanniomyces occi  
N;Alternate names: acid maltase; glucoamylase  
C;Species: Schwanniomyces occidentalis  
C;Date: 10-Mar-1994 #sequence\_revision 05-Apr-1995 #text\_change 18-Jun-1999  
C;Accession: JN0102  
R:Dohmen, R.J.; Strasser, A.W.M.; Dahlems, U.M.; Hollenberg, C.P.  
Gene 95, 111-121, 1990  
A;Title: Cloning of the Schwanniomyces occidentalis glucoamylase gene (GAM1) and its  
A;Reference number: JN0102; MUID:91071592; PMID:1979298  
A;Accession: JN0102  
A:Molecule type: mRNA  
A;Residues: 1-958 <DOH>  
A:Cross-references: GB:M60207; GB:M34666; NID:G169845; PIDN:AAA33923.1; PID:G169846  
C;Genetics:  
A:Gene: GAM1  
C:Superfamily: Schwanniomyces glucan 1,4-alpha-glucosidase GAM1; sucrose/isomaltase h  
F;1-22/Domain: signal sequence #status predicted <SIG>  
F;23-958/Product: Glucan 1,4-alpha-glucosidase #status predicted <MAT>  
F;156-876/Domain: sucrose/isomaltase homology <SIM>  
F;61,78,197,403,416,513,580,602,813,907/Binding site: carbohydrate (Asn) (covalent) #  
F;470/Active site: Asp #status predicted

Query Match 27.6%; Score 1279.5; DB 1; Length 958;  
Best Local Similarity 34.4%; Pred. No. 5.9e-89;  
Matches 315; Conservative 137; Mismatches 313; Indels 151; Gaps 27;

QY 30 AARTVLAVAVT-----MEGALRAEAATGSRSTGQVRLAVYASLETSLRLVRITDAD- 83  
DB 70 AARGDVLNVNTTPRGLTGILKKEATNIYDFOYLAFLV--EYQADTRLNVHIEPTDL 127

QY 84 HPRWEVPDIIIPR-APGDVLHDAPPASSAPLQGRVLSAGSDLVLTVHASPFRTVSR 142  
DB 128 SDVFLVPEHLVVKPLVEGD-----AQSYNFD-----NSDLVFEYSNTDFSEVIRS 173

QY 143 STGDTLFDTPAG-LVPRDKYLEVTSALPAGRASLYGLOB-----HTKSSFLRINDSFT 195  
DB 174 STKEVLFSTKGNPLVFESNQFIQFNGSLPKNHV-ITGLGESIHGLVNEPGSVK-----T 225

QY 196 LKNADIGASYDVNLVGHGPPFMDVR-APGTAGVYLLSSNMGMDVLYGGSVYTKVIGV 254  
DB 226 LFANDVQ-DPIDGNTYGVHPVYLDQRYDTETTHAYVWRTSAIQEVLIGEESTIHALSGV 284

QY 255 LDYFPFAGNPLAVVDQYTLIARPAHPYMGFPHQCRGYGLNVSDLRVAVAKARI 314  
DB 285 IDLYFPFSGTTPKDAIQOYVKEIGLPAFQYVSLGHQCRGWGDTIEKLSVVENFKNI 344

QY 315 PLEVMTDIDYMDGKDFLDRVNFATAELRPVDRLRHNAQYVLIIDPGIRV-DPIDA 373  
DB 345 PLETIWSIDIDYMDSYKDFYDPRFPLDEYRKFLDELHKNQHYVPIIDAAIYVNPENNA 404

QY 374 T---YGTFRVGHQDIFLKR-NGTNFVGNVMDGVYFDFPMHPAAAEFWARISLFRRTI 429  
DB 405 TDNEYQPFHYGNETDVLKNDPDSLYIGAVQVTLF-----SRFLSRKHSMDMKVI 455

```
QY 430 -----PVDGLWIDMNEISNF-----YNP----- 447
Db 456 KDWYELTPDGIWADWNEVSSFCVSCGCTGKYFENAYPPPTVGSKATSPYGVGFDVSNAS 515
QY 448 -----EPMNAL-----DDPPYRINNDGTGRPINNK 472
Db 516 EMKSIOSISATKTSSTSVSSSSSTIDYMTLAPCKGINPPYAIYMQGDSDLATH 575
QY 473 TVRPLAVHYGVTEYEHNLFGLLEBARATCGVLR-DTGRPRPVLRSRTEVSGRVTAYM 531
Db 576 AVSPNATHADGTVEYDHLNLYGLOENATYHALLEVPNKRPPMISRSTFPFRAGKWTGHM 635
QY 532 TGNAAATWGLRLYSINTMLSFGLFGNPMIGADICGNGNTTELCGRWTLQALGAFYFPSRD 591
Db 636 GGDNTADWAVAYFISIQAQSMGIAGLPFGGADVCGFNGNSDSELCSRWMLGSGFFPYFN 695
QY 592 HSAIFTVRLRYLWPSVAASGRKALGLRYQLLYFYFTLMAEAMTGAIPRLFFSYPHD 651
Db 696 HNYLGAIDOBPYWESVAEATRTSMAIRYLILLYLTHESHETTGLPILRAFSPWOPND 755
QY 652 VATYGVDRQPLLCRGVLVSPLEPGPTTVDAYFP-AGRWYRLDY--SLAVATRTGKHVR 708
Db 756 RSLUGVDNQFFVGDGLVTVLEFGVDKVGVPFGAGKEVYVDYTDQREVPFQGRNET 815
QY 709 LPAPADTVNHLTGGTILPQOSALTTSRARRTAFHLLVALAEDGTASGYLFLDDGDSPE 768
Db 816 LDAPLGHILPHIRGGNVLPQEPGYTVAESRQNPFGLLI VALDNDGKAQGSLYLDDGESLV 875
QY 769 YGRSDWSMVRFNKYKPNKGAIKVSEVVHNSYAQSRTLVILSKVVMGHRSPAAPKKLT 828
Db 876 VD-----SSLVFSVSDN-----TLSASPSGDYKADQPLANVTILGVGHK----PK--S 919
QY 829 VHVNSAEVEASSAGT 844
Db 920 VKFENANVDFTYKKST 935

RESULT 15
T10799
sucrose alpha-glucosidase (EC 3.2.1.48) / oligo-1,6-glucosidase (EC 3.2.1.10) - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: T10799
R:Chandrasena, G.; Osterholm, D.E.; Sunitha, I.; Henning, S.J.
Gene 150, 355-360, 1994
A:Title: Cloning and sequencing of a full-length rat sucrose-isomaltase-encoding cDNA.
A:Reference number: Z17155; MUID:95121929; PMID:7821806
A:Accession: T10799
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1841 <CHA>
A:Cross-references: EMBL:L25926; NID:9414818; PIDN:AAA65097.1; PID:g773669
A:Experimental source: strain Sprague-Dawley, intestine
C:Genetics:
A:Gene: SI
C:Function:
A:Description: sucrose alpha-glucosidase catalyzes the hydrolysis of the 1,2-alpha-D-glu
A:Pathway: carbohydrate digestion
C:Superfamily: sucrose/isomaltase; sucrose/isomaltase homology; trefoil homology
C:Keywords: carbohydrate digestion; glycosidase; hydrolase
F:199-846/Domain: sucrose/isomaltase homology <SIM>
F:937-983/Domain: trefoil homology <TRF>

Query Match 27.68; Score 1279; DB 2; Length 1841;
Best Local Similarity 36.08; Pred. No. 1.7e-88;
Matches 293; Conservative 140; Mismatches 285; Indels 96; Gaps 24;

QY 59 DVORLAVIASLETDSRLVRVITDADHPREVPQDIIPR-----PAPGDVLHDAPPASSAPL 114
Db 152 DIKSVLTTQTQGNRFRKLTIDPNKRYEVPHQVKEETGIPA-ADTLVD----- 201
QY 115 QGRVLPAGSDLVLTIVHASPFRTRRSRSTGDTLFTAPG-LVFRDKYLEVTSALPAGRA 173
```

Search completed: October 27, 2003, 10:22:01  
Job time : 29.9249 secs

```
Db 202 -----VQVSENPFSIKVIRKNNKVLCDISVGPLLYSNQVLOJSTRLPs--B 246
QY 174 SLYGLGEBHTKSPRLRHNDSTFLM---NADIGASYVDVNLVYSGHPFFYMDV-RAPGTAHGV 229
Db 247 YIYFGGHHK--RFRHDLVYKWTPIFRDBIPGDNHNLVYGHQTFEFGIGDTSKSYGV 304
QY 230 LLLSSNGMDV-LYSGSVYVYKVGVLDFYFAGPNPLAVVDQYLTQIARPAHPYMSFG 288
Db 305 FLMSNAMEVFIQPTIITYRTVTGGILDYFPLGDTPEOVVQQVQEVHWRPAMPAYNWL 364
QY 289 FHQCRYGYLANVSDLERVARYAKARIPILEVNMWTDIDYMDGKDFTLDRVNFNTAAELRPFV 348
Db 365 FOLSRWNTGSLDVTSEVVRNRREAGIPYDAQVTDIDYMEDHKEFTYDRVKFNG-LPEFA 422
QY 349 DRLHNAQKYYLILDPGIRVD--PIDATYGTFFVGMQODIFLKRNG--TNFVGNVWPGDV 404
Db 423 QDLH-NHOKYIIILDPALISINKRANGAEYQTYVRGNEKNVWVNESDGTPLIGVWPGLT 481
QY 405 YPPDFMHPAAAEFWAREISLFRRTIPVDGLWIDMNEISNF-----YNPE 448
Db 482 VYPOFTNPQTIEWANECNLFHQVEYDGLWIDMNEVSSFIQGSINLKGVLILVNLVPPF 541
QY 449 PMNALDDPPYRINNDGTGRPINNKTVRPLAVHYGVTEYEHNLFGLLEARATGRGVLR- 507
Db 542 TPGILDVKMY-----SKTLCMDAVQHWG-KQYDVHSLYGYGMAIATEQAVERV 588
QY 508 DTGRPPFVLSRSTFVSGRYTAYMTGDNAATWGLRLYSINTMLSFGLGMPMIGADICGF 567
Db 589 FPNKRSFILTRSTGGSRHANHLWLDNTASWQNEWSITOMLEFGIFGMLVLGATSCGF 648
QY 568 NGNTTEELCGRWIQLGAFYPFESRDHSALFTVRRRELYLWPSVA---ASGRKALGLRYQL 623
Db 649 LADTTEELCRWMLQALGAFYFESRNHNA-----EGYMEQDPAYFGQDSSRHYLTIRYTL 702
QY 624 PYFTYLTAEAMTGAIPRLFFSYPHDVATYGVDRQPLLCRGVLVSPVLEPGPTTVDAY 683
Db 703 PFLYTLFVRAHMFGETVARPFLYFEDTNSWIEDTQFLMGPAALLITPVLPRGVENVISAY 762
QY 684 FPAGRWYLYDYSLAVATRTGK-HVRLPAPADTVNHLTGGTILPQOSALTTSRART 742
Db 763 IPNATM---YDYETGIRKPRWRKERINMYLPQDKIGLHRLGGYIITPOEPDVTITASRKNP 819
QY 743 FHLVALAEDGTASGYLFLDDGDSPEYGRSDWSMVRFNKYKPNKGAIKVSEVVHNSY 802
Db 820 LGLIVALDDNQAAKGELEFMDGSKDSIEKKMY--ILYTPSVSNNEVLV----NCHSSY 873
QY 803 AQSRTLVISKVVLGMHRSPAAPKKLTIVHNSAEV 836
Db 874 AEGTSLAFKTIKVLGLREDV--RSITVGENDQQM 905
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:20:43 ; Search time 105.357 Seconds  
(without alignments)  
1393.955 Million cell updates/sec

Title: US-10-043-418-1

Perfect score: 4630

Sequence: 1 MATGVLLCLCLCFAPRL.....IGGLSIVGVGEPELVKVMYS 877

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 629382 seqs, 167460630 residues

Total number of hits satisfying chosen parameters: 629382

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US05\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4630	100.0	877	14	US-10-043-418-1
2	2451	52.9	902	14	US-10-043-418-4
3	2422.5	52.3	903	14	US-10-043-418-3
4	2370	51.2	914	14	US-10-043-418-2
5	1924	41.6	707	12	US-10-032-189-125
6	1329.5	28.7	1827	9	US-09-819-247-2
7	956	20.6	912	12	US-10-032-189-44
8	953.5	20.6	914	12	US-10-032-189-40
9	952.5	20.6	914	12	US-10-032-189-42
10	913.5	19.7	966	12	US-10-032-189-120
11	911.5	19.7	943	12	US-10-032-189-123
12	911.5	19.7	944	12	US-10-032-189-122
13	903.5	19.5	966	12	US-10-032-189-121
14	900	19.4	967	12	US-10-032-189-38
15	878.5	19.0	944	12	US-10-032-189-124

16	818	17.7	565	15	US-10-102-806-557	Sequence 557, Appl
17	679.5	14.7	693	12	US-10-228-063-5	Sequence 5, Appl
18	679.5	14.7	712	12	US-10-228-063-27	Sequence 27, Appl
19	679.5	14.7	718	12	US-10-228-063-26	Sequence 26, Appl
20	679.5	14.7	718	12	US-10-228-063-36	Sequence 36, Appl
21	646	14.0	788	15	US-10-156-761-14497	Sequence 14497, A
22	541.5	11.7	235	9	US-09-734-569-132	Sequence 132, App
23	466	10.1	1066	10	US-09-280-197-5	Sequence 5, Appl
24	466	10.1	1066	10	US-09-423-126-3	Sequence 3, Appl
25	464.5	10.0	1070	10	US-09-280-197-6	Sequence 6, Appl
26	464.5	10.0	1070	10	US-09-423-126-4	Sequence 4, Appl
27	426.5	9.2	642	15	US-10-156-761-10107	Sequence 10107, A
28	385.5	8.3	1092	10	US-09-423-126-5	Sequence 5, Appl
29	385	8.3	188	9	US-09-734-569-134	Sequence 134, App
30	355	7.7	1091	10	US-09-280-197-2	Sequence 2, Appl
31	355	7.7	1091	10	US-09-423-126-2	Sequence 2, Appl
32	354	7.6	1088	10	US-09-280-197-1	Sequence 1, Appl
33	354	7.6	1088	10	US-09-423-126-1	Sequence 1, Appl
34	279	6.0	570	10	US-09-423-126-6	Sequence 6, Appl
35	277.5	6.0	199	9	US-09-734-569-130	Sequence 130, App
36	138	3.0	3686	15	US-10-156-761-11179	Sequence 11179, A
37	135	2.9	1038	15	US-10-156-761-13515	Sequence 13515, A
38	122.5	2.6	6396	12	US-09-940-316B-72	Sequence 72, Appl
39	119	2.6	613	12	US-10-120-145-5	Sequence 5, Appl
40	119	2.6	722	15	US-10-136-841-6	Sequence 6, Appl
41	116	2.5	3816	11	US-09-808-880-3	Sequence 3, Appl
42	114	2.5	547	11	US-09-943-857-2	Sequence 2, Appl
43	111.5	2.4	317	10	US-09-280-197-9	Sequence 9, Appl
44	111	2.4	1578	12	US-09-940-316B-31	Sequence 31, Appl
45	110.5	2.4	1463	10	US-09-971-536-69	Sequence 69, Appl

ALIGNMENTS

RESULT 1  
US-10-043-418-1  
; Sequence 1, Application US/10043418  
; Publication No. US20020184662A1  
; GENERAL INFORMATION:  
; APPLICANT: Henson, Cynthia A.  
; APPLICANT: Muslin, Elizabeth H.  
; APPLICANT: Clark, Suzanne E.  
; TITLE OF INVENTION: Modified barley alpha-glucosidase  
; FILE REFERENCE: 960296.97486  
; CURRENT APPLICATION NUMBER: US/10/043.418  
; PRIOR FILING DATE: 2002-06-25  
; PRIOR APPLICATION NUMBER: 60/260,787  
; PRIOR FILING DATE: 2001-01-10  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 877  
; TYPE: PRT  
; ORGANISM: Barley  
US-10-043-418-1

Query Match	100.0%;	Score	4630;	DB	14;	Length	877;
Best Local Similarity	100.0%;	Pred. No.	0;				
Matches	877;	Conservative	0;	Mismatches	0;	Indels	0;
						Gaps	0;
Qy	1	MATGVLLCLCLCFAPRLCSKEEGPLAARTVLA	VAVTM	EGALRAEATGGRSS	TDV	60	
Db	1	MATGVLLCLCLCFAPRLCSKEEGPLAARTVLA	VAVTM	EGALRAEATGGRSS	TDV	60	
Qy	61	QLAVYASLETDSRLVRITDADHPRWEVQDII	PRPAGD	VLHDAPPASSAP	LQGRVLS	120	
Db	61	QLAVYASLETDSRLVRITDADHPRWEVQDII	PRPAGD	VLHDAPPASSAP	LQGRVLS	120	
Qy	121	PAGSDLVLTVHASPRFTVSRRTSGDTLFT	OTAPGL	VRDKYLEVTSALP	AGRASLYGLGE	180	
Db	121	PAGSDLVLTVHASPRFTVSRRTSGDTLFT	OTAPGL	VRDKYLEVTSALP	AGRASLYGLGE	180	



QY 805 SR:LVISKVVLGMGHRSPAAPKLTTHVNSAEVEASSAGTRYQNAGGLGGVAHIGGLSLV 864

```
QY 566 GFGNNTTEELCGEWIOLGAPYFSDHSAIFTVRRELYLWPSVAASCRKALGLRYOLLPY 625
Db 600 GFAESTTEELCCRWIQGAFYFSDHSAITTHQELYLWESVAASARTVLGJRYELLPY 659
QY 626 FYTLMYEAHMTGAPARLPFFSYPHDVATYGVDRQFLGRVLVSPVLEPGPTTVDAYPF 685
Db 660 YYTLMYDANLRGSIARPLSFTFPDDVATYGISQFLIGIIVSPVLOFGSSIVNAYSP 719
QY 686 AGRWRYLYDSLAVATGKXRLPAPADTVNVHLTGGTILPLQOQSALTTSRRARTAFHL 745
Db 720 RGNWVLSNYSVSAGTVLSAPPDHNVIHHEGNIVAMQGEAMTTOAARSTPFHL 779
QY 746 LVALAEDGTASGYLFLDDGDSPEY-GRSDWSMVRENYKIPNNKGAIKVKSEVHNSYAO 804
Db 780 LVNMSHVASTGELFLDNGIEMDGGPGKRWTLVRPFAESGINN--LTISSEVNRGYAM 837
QY 805 SRTLISKVLMGHRSPAAPKLLTVHNSAEVEASS-SAGTRYQAGGLGGVAHIGLGLSL 863
Db 838 SQRWMBKITILGKRRVKIKXYTQKDAGAIKVKGLGRITSSHNQGGF-FVSVISDLRQ 896
QY 864 VUGEPELVK 873
Db 897 LVGQAFKLEL 906

RESULT 5
US-10-032-189-125
; Sequence 125, Application US/10032189
; Publication No. US20030170630A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook II, John P
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Zerrhusen, Bryan D
; APPLICANT: Patturajan, Meera
; APPLICANT: Grosse, William M
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Shimkets, Richard A
; APPLICANT: Grosse, William M
; APPLICANT: Szekeres, Edward S
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Li, Li
; APPLICANT: Casman, Stacie J
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Gorman, Linda
; APPLICANT: Gangolli, Esha A
; APPLICANT: Fernandes, Elma R
; APPLICANT: Rieger, Daniel K
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Gunther, Erik
; APPLICANT: Millet, Isabelle
; APPLICANT: Sciore, Paul
; APPLICANT: Ellerman, Karen
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glenna
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; CURRENT APPLICATION NUMBER: US/10/032,189
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,495
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/258,171
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/269,940
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 60/274,192
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/277,826
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/279,840
```

```
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/282,981
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/283,656
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/309,247
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/311,754
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/313,331
; PRIOR FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 260
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 125
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Glycobyl
; OTHER INFORMATION: hydrolases family 31
US-10-032-189-125

Query Match 41.6%; Score 1924; DB 12; Length 707;
Best Local Similarity 54.1%; Pred. No. 9e-176;
Matches 392; Conservative 112; Mismatches 188; Indels 32; Gaps 19;

QY 108 PASSAPLQGRVLSPAGSDLVLTVHAS-PFRFTVSRSTGDTLPTAPG-LVFRDKYLEVT 165
Db 5 PASS-----SPABSDLYDLVLSNCGPFGFEVIRKSTGDLVPTDTFGPLVFSDFQLQLS 56
QY 166 SALPAGRASLYGLGEHTKSSFRLRHNDSPFTLWNADIGASYVDVNLVYGHSHPPYMDVRAPGT 225
Db 57 TSLPS--EYIVGLGEHAHKLFRDNTETTYTLWNRDVGPYSGDNLYGSHPPYMLEDSGN 114
QY 226 AHGVLLSSNGMDVLYG-GSYTVYKVIIGVLDYFPPFAGPNPLAVVDYQTQIARAPAMPY 284
Db 115 AHGVFLNSNAMEVDIGPGPALYRVVIGGILDFYFFLGGTPEDLVQQTTELGIRPALPPY 174
QY 285 WSPGHQCRVYGLVNSDLERVARVAKARIPLEVYMTWTDIDYMDGFKOFTLDRVNFIAEL 344
Db 175 WSLGFHLCRWGTYNVSEKTVVDCGRKANIPLDQWLDIDYMDGXKFTWDFVFRPPGE- 233
QY 345 RPFVDRLHRNAQKYVLIILDPGIRVDPIDATYGTFRGMOQDIFLKR-NGTNFVGNVWPGD 403
Db 234 -DFVKKLHAKGKYVLIILDPALISVD--SASYYPYERGKEKGVFKVKNPGSDYICEVMPGY 290
QY 404 VYFPDPMHAAAEFAWEISIFRPTIPVDGLWIDMNEISNFPYNPEPMA-LDDPPYRINN 462
Db 291 TAFPDFTNPEARKWMADEIKDFHDSLPFDGIWIDMNEPSESEPCPNDSNLNYPY-APN 349
QY 463 DGTGRPINNKTVRPLAVHYGVVTEYEEHNLFCLEARATGRGLRDT-GRPPFVLSRSTF 521
Db 350 DGDG-PLSSKTMCDADVHYGVVEHYDVNLYGLSEAKATYEAALKKVTGGKAPFVLSRSTF 408
QY 522 VSGRYTAYTGDNAATWMDLRYISINTWLSFGLGPMPIGADI CGFNNGNTTEELCGRWIQ 581
Db 409 AGSGRYAGHWTDNTASWDDLKYSIPGVLSFNLGIPFVGADICGFNGNTTEELCVRWMQ 468
QY 582 LGAFYFSDHSAIFTVRRELYLWPSVAA-SGRKALGLRYOLLPYFYTLMYEAHMTGAPI 640
Db 469 LGAFYFSDHSAIFTVRRELYLWPSVAA-SGRKALGLRYOLLPYFYTLMYEAHMTGAPI 640
QY 641 APPLFFSYPHDVATYGVDRQFLGRVLVSPVLEPGPTTVDAYPFAGRWLYDYSLAVA 700
Db 528 APPLFFSYPHDVATYGVDRQFLGRVLVSPVLEPGPTTVDAYPFAGRWLYDYSLAVA 700
QY 529 WRPJFFEFDDAETDYDIDRQFLWGSALLVAPVLEPGATSVKAYLPGGRWYDLYT-GAGEA 587
QY 701 TRTGKVELPAPADTVNVHLTGGTILPLQOQSALTTSRRARTAFHLVALAEDGTASGYLF 760
Db 598 SR-GGNVTLSPALDKIPVHVVGSGSIIPTEOPALTTTESRDNPFLHLLVALDDNGTASGELY 646
QY 761 LDDGDSPEYGRSDWSMVRENYKIPNNKGAIKVKSEVHNSYAO SRTLISKVLMGHRSPA 819
Db 647 LDDGESID-TQGDYLLVQFS---ANNN--TLTGTGVVTVTYGKNSNTLLEKITILGVGN 700
```



QY	820	SPAA 823		633	AHMTGAPIARELFFSYPHDVATYGVDRQFLGRGVLSVPLEPGTPTVDAYFPAGRWRL	692
Db	701	EPAA 704		706	ARHGETVARVPLHEFYEDTNSMTEDTEFLWGPALLITPVLKQAGADTVSAYIPDAIM---	762
RESULT 6						
US-09-819-247-2						
; Sequence 2, Application US/09819247						
; Patent No. US2001003663SA1						
; GENERAL INFORMATION:						
; APPLICANT: Waldman, Scott A.						
; APPLICANT: Park, Jason						
; APPLICANT: Schulz, Stephanie						
; TITLE OF INVENTION: Compositions And Methods For Identifying And Targeting Cancer Cell						
; TITLE OF INVENTION: Alimentary Canal Origin						
; FILE REFERENCE: T002413						
; CURRENT APPLICATION NUMBER: US/09/819,247						
; PRIOR FILING DATE: 2001-03-27						
; PRIOR FILING DATE: 2001-03-27						
; PRIOR FILING DATE: 2000-03-27						
; NUMBER OF SEQ ID NOS: 2						
; SOFTWARE: PatentIn version 3.0						
; SEQ ID NO 2						
; LENGTH: 1827						
; TYPE: PRT						
; ORGANISM: Homo sapiens						
US-09-819-247-2						
Query Match 28.7%; Score 1329.5; DB 9; Length 1827;						
Best Local Similarity 36.4%; Pred. No. 7.3e-118;						
Matches 307; Conservative 144; Mismatches 309; Indels 83; Gaps 23;						
QY	59	DVORLAVYASLETSRLVRITDADHPREVPDIIER---	PAGDVLHDAPPASSAPIQ 115	693	YDY-SLAVATRTGXHRLPAPADTVNVHLTGGTILPLQOQSALTTSRRRTAFHLLVALAE	751
Db	142	DINSVLTQNTQPNRFRKITDNNRRYEPVHQYKFTGPTVSOTLYD-----	191	706	ARHGETVARVPLHEFYEDTNSMTEDTEFLWGPALLITPVLKQAGADTVSAYIPDAIM---	762
QY	116	GRVLSFAGSDVLTVTHVASPPFRFTVSRRSTGDTLFTDAPG-LVFRDKYLEVTISALPAGRAS	174	693	YDY-SLAVATRTGXHRLPAPADTVNVHLTGGTILPLQOQSALTTSRRRTAFHLLVALAE	751
Db	192	-----VKVACNPESIQVIRKNGKTLFDTSIGPLVYSDQYLQISARLPD--Y	237	706	ARHGETVARVPLHEFYEDTNSMTEDTEFLWGPALLITPVLKQAGADTVSAYIPDAIM---	762
QY	175	LYGLGHTKSSFLRINDSTLM---NADIGASYVDVNLGSHPPFMDVR-APGTAHGLV	230	693	YDY-SLAVATRTGXHRLPAPADTVNVHLTGGTILPLQOQSALTTSRRRTAFHLLVALAE	751
Db	238	IYIGIEQVHK--RFRHLSKTKPIETFDOLPQDNNNNLYGHQTFEMCIEDTSKSFVGF	295	706	ARHGETVARVPLHEFYEDTNSMTEDTEFLWGPALLITPVLKQAGADTVSAYIPDAIM---	762
QY	231	LLSSNGMDV-LYGSVYTVKVGVLDFEYFAGPNPLAVVDQYQTOLARPAHPYMSFCF	289	693	YDY-SLAVATRTGXHRLPAPADTVNVHLTGGTILPLQOQSALTTSRRRTAFHLLVALAE	751
Db	296	LMSNAMEIFIQTPITVTRVGTGIIQYFILLGTPEQVVQYQQVGLPAPMPAYWNLGF	355	706	ARHGETVARVPLHEFYEDTNSMTEDTEFLWGPALLITPVLKQAGADTVSAYIPDAIM---	762
QY	290	HOCRYGLNVSDLERVARYAKARIFLEVNMWTDIDYMGFKDFTLDRVNFRTAAELRPFVD	349	693	YDY-SLAVATRTGXHRLPAPADTVNVHLTGGTILPLQOQSALTTSRRRTAFHLLVALAE	751
Db	356	QLSRWNYKSLDVVKEVVRNRREAGIPFDQVTDIDYMEDKKDFTYDQVAFNG-LPQFVQ	413	706	ARHGETVARVPLHEFYEDTNSMTEDTEFLWGPALLITPVLKQAGADTVSAYIPDAIM---	762
QY	350	RLHNAQKYLILDPGRVD--PIDATYGTFFVGMQODIFLKEN--GTNFFVGNVWPGDY	405	693	YDY-SLAVATRTGXHRLPAPADTVNVHLTGGTILPLQOQSALTTSRRRTAFHLLVALAE	751
Db	414	LDHGGQKYLILDPGRVD--PIDATYGTFFVGMQODIFLKEN--GTNFFVGNVWPGDY	405	706	ARHGETVARVPLHEFYEDTNSMTEDTEFLWGPALLITPVLKQAGADTVSAYIPDAIM---	762
QY	406	FPDFMHPAAEFWAREISLPRITPVDFGLMDNNEISNFYNPEP-----MNALDDPPYRN	461	693	YDY-SLAVATRTGXHRLPAPADTVNVHLTGGTILPLQOQSALTTSRRRTAFHLLVALAE	751
Db	474	YPDFTNPNCDWANECSIFHOEVQYDGLWDMNEVSSFIQSGTKGVNKNLYPPF--T	531	706	ARHGETVARVPLHEFYEDTNSMTEDTEFLWGPALLITPVLKQAGADTVSAYIPDAIM---	762
QY	462	NDGTGRPINNKTVPRLAVHYGVTEYEEHNLFGLEEARATGROVLR-DTGRRPFLVSRST	520	693	YDY-SLAVATRTGXHRLPAPADTVNVHLTGGTILPLQOQSALTTSRRRTAFHLLVALAE	751
Db	532	PDILDKLWYKTIKMDAVQWNG-KQYDVHSLYGSMAIATEQAVQVFNKRSFILTRST	590	706	ARHGETVARVPLHEFYEDTNSMTEDTEFLWGPALLITPVLKQAGADTVSAYIPDAIM---	762
QY	521	FVSGGRVTAWTGDNATWGLDLYSINTMLSGFLGPMPLGADICGFGNGTTEELCGRW	580	693	YDY-SLAVATRTGXHRLPAPADTVNVHLTGGTILPLQOQSALTTSRRRTAFHLLVALAE	751
Db	591	FAGSGRAAHWLGNDNTASWEQMEWSITGMLFSLGIPLVGADICGFAETTEELCRWM	650	706	ARHGETVARVPLHEFYEDTNSMTEDTEFLWGPALLITPVLKQAGADTVSAYIPDAIM---	762
QY	581	QLGAFYFPRSDHSA-----IFTVRRELYLWPSVAASGRKALGRLYQLLPVFTLWYE	632	693	YDY-SLAVATRTGXHRLPAPADTVNVHLTGGTILPLQOQSALTTSRRRTAFHLLVALAE	751
Db	651	QLGAFYFPRSDHSA-----IFTVRRELYLWPSVAASGRKALGRLYQLLPVFTLWYE	632	706	ARHGETVARVPLHEFYEDTNSMTEDTEFLWGPALLITPVLKQAGADTVSAYIPDAIM---	762

```
/ PRIOR APPLICATION NUMBER: 60/309,247
/ PRIOR FILING DATE: 2001-07-31
/ PRIOR APPLICATION NUMBER: 60/311,754
/ PRIOR FILING DATE: 2001-08-17
/ PRIOR APPLICATION NUMBER: 60/313,331
/ PRIOR FILING DATE: 2001-08-17
/ NUMBER OF SEQ ID NOS: 260
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 44
/ LENGTH: 912
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-032-189-44

Query Match      20.6%; Score 956; DB 12; Length 912;
Best Local Similarity 28.3%; Pred. No. 2.1e-82;
Matches 267; Conservative 165; Mismatches 336; Indels 174; Gaps 31;

QY 32 RTVLAVAVTMEGALRAEAATGRRSTGQVQRLAVVAVSLETD-SRLRVRIITDADHPREVP 90
DB 44 RALLDSVTIDSDTRFQII-----NEASKVPLLAIEYIGEGIFRLKINEETPLKPRFEVP 99
QY 91 QDIIPRPAQDVLHDAPPASSAPLQGRVLSFAGSDLVLTVHASPPR-----FTVSRR 142
DB 100 DVLTSKPSTVRLI-----SCSGDTGSLILADGKGLKCHITANFPKVDLVSEEEWISIN 154
QY 143 STG-----DTLFDTPA-----GLVFR--DKLEVTSAIPA--- 170
DB 155 SLGQLYFEHLQILHKQRAAKENEETSVDTSQENQEDLGLWEEKFEGFVDIKANGPSSIG 214
QY 171 -----GRASLYGLGEHTFKSSFLRHNDSTFLWADLGA--SYVDVNLVGLSHPPYMDVRA 222
DB 215 LDPSLHGFEHLVGPQHAESHQLKNTGDARLYNLVDYGVQIDKMGIIYSVP-YLLAHK 273
QY 223 PGTAGVLLLSN-----GMDVLYGGSYVYTKVIGVGLDIFYF 259
DB 274 LGRTIGFNLNASETLVBINTEPAVEYTLTQMGFVAAKQKVRSTRVHWMSESGLIDVFL 333
QY 260 FAGNPPLAVDQYTLIARPAAMPVWSPGPHQCRQYGLNVSLDLERVVARYAKARIPLEV 319
DB 334 LTGPTPSPVFKQYSHLTQTQMPPLFSLGYHQCWRNYEDEQDVKAVDAGDEHDIPYDAM 393
QY 320 WTDIDYMGDFKDLDRVNFVTAELRPVDFVDRHNAQKYLILDPGIRVDPIDATYTFV 379
DB 394 WLDIHTEGKRYFTWDKNRF--PNPKRMQELLRSKKRLVVISDPHIKIEP--DYSSVYV 448
QY 380 RGMQODIFLK-RNGTNFVGNVWPGDVPFDPFMHFAAAEFWAREISLPRPIPV-DG---- 433
DB 449 KAKDQGFVKNQGBEDDFEGVCWPGLSYLDFTNPKVREWYS---SLF--APFVYQGSTOI 503
QY 434 --LWIDMNEISNFYNPEPMNALDDPPYRINNDGTGRPINNKTVRPLAVHYGGVTEYEHN 491
DB 504 LFLMNDMNEPSVFRGPE-----CTMKNAIHGNWEHRELHN 540
QY 492 LFGLEAARATGRVUL-RTGR-RPVLRSRTFVSGRYTATWTGDNAAATWGDRLYS:NTM 549
DB 541 IYGFTHOMATAEGLIKRSGKERPFVLTSPFAGSQYKQYCAVWTGNTAENSNLKISIPML 600
QY 550 LSPGLFGMPMLGADICGNGNNTTELGRWTLQCAFYFPFSDHSAIETVBRRELYLWESVA 609
DB 601 LTLSTIGISFCGADIGGFIQNPETELLVWVTAQAGAYQFFRGHAYMTKREPWLFI---- 656
QY 610 ASG-----RKALRLRYQLLPYFTLLMYEAHMTGAP:ARLPFFSYPHDVATYGVDRQFL 662
DB 657 --GEHTRLIREAIRERYGLLPYWSLYFAHVASQPMRPLWVEFPDELKTFDMEDEYM 714
QY 663 LGRGVLSVPVLEPGTTVDATFPAGR--WYRLYOYSLAVATRTGKVRLLPAPADTVNVHL 720
DB 715 LGSALLVHPVTEPKATVDVLEPLGNSWYV---YDKTFPAHWEGGCTVKIPVALDTPVFQ 771
QY 721 TGGTILPLQQS-ALTTSRARATFHLVALAEDCTASCYLFLLDDGSDPEYCRSRDWSMVR 779
DB 772 RGGSVIPKTTVGKSTGMWTSYGLRVALSTQGSVSGELVLDGDSFGYLHQKQFLHRK 831
```

```
QY 780 FNYKIPNNKGAIKVKSEVYVHNSYAQSR-----TLVISKVLMGHRSPAAPKKLTWHVNSA 834
DB 832 FSF-----CSSVLINSFADQQRGHYPSKCVKEKILVLGFRK--EPSSVTHSSDG 878
QY 835 E---VEASSSAGTRYQNAAGLGGVGAHIGGLSLVVGEEPELVK 873
DB 879 KDQPVAFYTCAT-----SILSLEKLSLNIAITDMEVRI 911
```

## RESULT 8

```
US-10-032-189-40
/ Sequence 40, Application US/10032189
/ Publication No. US20030170630A1
/ GENERAL INFORMATION:
/ APPLICANT: Alsobrook II, John P
/ APPLICANT: Tchernev, Velizar T
/ APPLICANT: Liu, Xiaohong
/ APPLICANT: Spytek, Kimberly A
/ APPLICANT: Zernhusen, Bryan D
/ APPLICANT: Patturajan, Meera
/ APPLICANT: Grosse, William M
/ APPLICANT: Lepley, Denise M
/ APPLICANT: Burgess, Catherine B
/ APPLICANT: Shinkets, Richard A
/ APPLICANT: Grosse, William M
/ APPLICANT: Szekeres, Edward S
/ APPLICANT: Vernet, Corine A.M.
/ APPLICANT: Li, Li
/ APPLICANT: Casman, Stacie J
/ APPLICANT: Boldog, Perenc L
/ APPLICANT: Gorman, Linda
/ APPLICANT: Gangolli, Bsha A
/ APPLICANT: Fernandes, Rima R
/ APPLICANT: Rieger, Daniel K
/ APPLICANT: Edinger, Shlomit R
/ APPLICANT: Gunther, Erik
/ APPLICANT: Millet, Isabelle
/ APPLICANT: Sciore, Paul
/ APPLICANT: Ellerman, Karen
/ APPLICANT: MacDougall, John R
/ APPLICANT: Smithson, Glenna
/ TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
/ FILE REFERENCE: 21402-228
/ CURRENT APPLICATION NUMBER: US/10/032,189
/ CURRENT FILING DATE: 2001-12-21
/ PRIOR APPLICATION NUMBER: 60/257,495
/ PRIOR FILING DATE: 2000-12-21
/ PRIOR APPLICATION NUMBER: 60/258,171
/ PRIOR FILING DATE: 2000-12-20
/ PRIOR APPLICATION NUMBER: 60/269,940
/ PRIOR FILING DATE: 2001-02-20
/ PRIOR APPLICATION NUMBER: 60/274,192
/ PRIOR FILING DATE: 2001-03-08
/ PRIOR APPLICATION NUMBER: 60/277,826
/ PRIOR FILING DATE: 2001-03-22
/ PRIOR APPLICATION NUMBER: 60/279,840
/ PRIOR FILING DATE: 2001-03-29
/ PRIOR APPLICATION NUMBER: 60/282,981
/ PRIOR FILING DATE: 2001-04-11
/ PRIOR APPLICATION NUMBER: 60/283,656
/ PRIOR FILING DATE: 2001-04-13
/ PRIOR APPLICATION NUMBER: 60/309,247
/ PRIOR FILING DATE: 2001-07-31
/ PRIOR APPLICATION NUMBER: 60/311,754
/ PRIOR FILING DATE: 2001-08-17
/ PRIOR APPLICATION NUMBER: 60/313,331
/ PRIOR FILING DATE: 2001-08-17
/ NUMBER OF SEQ ID NOS: 260
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 40
/ LENGTH: 914
/ TYPE: PRT
```

; ORGANISM: Homo sapiens	
US-10-032-189-40	
Query Match 20.6%; Score 953.5; DB 12; Length 914;	
Best Local Similarity 28.8%; Pred. No. 3.7e-82;	
Matches 260; Conservative 161; Mismatches 308; Indels 173; Gaps 31;	
Qy 74	RLVRITDADHPRWEVDIIPAPAGDVLDHAPPASSAPLQGRVLSPPAGSDLVLTVHAS 133
Db 83	RLKINEETPLKPRFEVDVLTSPSTVRLI-----SCSGDTGSLIADGKDLKCHITAN 137
Qy 134	PER-----FTVSRSTG-----DTLPTAP-----GLVFR 158
Db 138	PFKVDLSEEVVISINSLQLYFEHLIQLHKORAAKENEETSVDTSQENQEDLGLWEE 197
Qy 159	--DRKLEVTGALPA-----GRASLYGLGHRTKSSFLRH---NDSFTLWNAIDGA- 203
Db 198	KFGKFVDIKANGFSSIGLDFSLGHFEHLYGIPQHAEE-SHQLKNTGDDAYRLYNLDVGY 256
Qy 204	-SYVDNLYGSHRPFYMDVRAPGTAHGLVLLSSN-----GMDV 239
Db 257	QIDYDMGIYGVSP-YLAHLKLGRTIGIPWLNASSETLVEINTEPAVEYTLTQMGFVAAKOK 315
Qy 240	LYGGSYVTVYVIGVGLDFYFFAGNPLAVVDQYTLIARPAHPMPYVSFGHQCRYGILNV 299
Db 316	VRSRTHVHMSGSIIDVFLLTGTPSDVFKQYSHLTGTQAMPPLFSLGVHQCRRWYDE 375
Qy 300	SDLERVARAKARIPLEVQMTDIDYMDGFKDTLDRVFTAEALPFFVDRLHRNAQKYV 359
Db 376	QVKAVDAGFDEHDIPYDAMWMLDIEHTGKRYFTWQKRF--PNPKRMQELLRSKKRKLIV 433
Qy 360	LILDPGRVRPIDATYGTFRGMQODIFLK-RNNTNFVGNVMPGDVYFDFPMHAAAEFW 418
Db 434	VISDPHIKIDP---DYSVYKAKDQGFVKNQEGEDEFVCMFGLSSYLDFTNPKVREMY 490
Qy 419	AREISLFRRTIPV-DG-----LWIDMNEISNFYNPEPMALDDPPYRINNDOTGRPINN 471
Db 491	S---SLF--APPVYQGSTDILFLWMDNNEPSVFRGPE-----522
Qy 472	KTVRPLAVHYGGVTEYEERHNLFGLEARATGRGVL-RDTGR-RPFVLSRSTFVSGRYTA 529
Db 523	QTMQNAIHGNEHRELRNIYGYHOMATAEGLIKSKKERPFVLTFRFFAGSOKYGA 582
Qy 530	YWTGDNAATWGLRYSINTMLSFGLFQMPMIGADICGFNGNTTEELCGRMQLGAPYFYS 589
Db 583	VMTGDNTAENSNLKISIPMLLTLSITGIFSGADIGGFIGNPETELLVRWYQAGAYQPF 642
Qy 590	RHSAITVTRRELYLWPSVAASG-----RKALGLRYQLLPFYTLMYEAMTGAPIAR 642
Db 643	RGHATMNTKREPLWF-----GEEHRLIREAIRERYGLPYWLSLYFAHVAASQFVMR 696
Qy 643	PLFFSYPHDVATYGVDRQFLILGRGVLSVPLEPGFTTVDAYFPAGR--WYRLDYDSLAVA 700
Db 697	PLWVEFDDELKTDFMEDEYMLGSALLVHPVTEPKATTVDFLPSNEVW---YDKTFPAH 753
Qy 701	TRTKHVRLPAPATVNVHLTGTILPQQOS-ALTTSRARRTAPHLVALAEDGTASGYL 759
Db 754	WEGGCTVKIPVALDTIPVFORGGSVIPKITYGKSTGMWTESSYGLRVALSTKGSSVGL 813
Qy 760	FLDDGDSPEXGRRSDMSVRENYKIPNNKGAIKYKSEVHNSVAQSR-----TLVLSKV 814
Db 814	YLDGHSFYQLHOKQFLHRKFSF-----CSSVLINSFADQRGHYPSKCVVEKIL 862
Qy 815	LMGHRSPAPKKLTVHNSAS---VEASSSAGTRYQAGGLGGVAHITGGLSLVVGBEFEL 871
Db 863	VLGFRK--EPSSVTTHSDGDKQDPVATYCAKT-----STLSLEKLSLIATDMEV 911
Qy 872	KV 873
Db 912	RI 913
RESULT 9	

US-10-032-189-42	
; Sequence 42, Application US/10032189	
; Publication No. US20030170630A1	
; GENERAL INFORMATION:	
; APPLICANT: Alsobrook II, John P	
; APPLICANT: Tchernev, Velizar T	
; APPLICANT: Liu, Xiaohong	
; APPLICANT: Spytek, Kimberly A	
; APPLICANT: Zehuseen, Bryan D	
; APPLICANT: Patturajan, Meera	
; APPLICANT: Grosse, William M	
; APPLICANT: Lepley, Denise M	
; APPLICANT: Burgess, Catherine E	
; APPLICANT: Shimkets, Richard A	
; APPLICANT: Grosse, William M	
; APPLICANT: Szekeres, Edward S	
; APPLICANT: Vernet, Corine A.M.	
; APPLICANT: Li, Li	
; APPLICANT: Casman, Stacie J	
; APPLICANT: Boldog, Ferenc L	
; APPLICANT: Gorman, Linda	
; APPLICANT: Gangolli, Esha A	
; APPLICANT: Fernandes, Elma R	
; APPLICANT: Rieger, Daniel K	
; APPLICANT: Edinger, Shlomit R	
; APPLICANT: Gunther, Erik	
; APPLICANT: Millet, Isabelle	
; APPLICANT: Sciore, Paul	
; APPLICANT: Ellerman, Karen	
; APPLICANT: MacDougall, John R	
; APPLICANT: Smithson, Glennda	
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same	
; FILE REFERENCE: 21402-228	
; CURRENT APPLICATION NUMBER: US/10/032,189	
; CURRENT FILING DATE: 2001-12-21	
; PRIOR APPLICATION NUMBER: 60/257,495	
; PRIOR FILING DATE: 2000-12-21	
; PRIOR APPLICATION NUMBER: 60/258,171	
; PRIOR FILING DATE: 2000-12-20	
; PRIOR APPLICATION NUMBER: 60/269,940	
; PRIOR FILING DATE: 2001-02-20	
; PRIOR APPLICATION NUMBER: 60/274,192	
; PRIOR FILING DATE: 2001-03-08	
; PRIOR APPLICATION NUMBER: 60/277,826	
; PRIOR FILING DATE: 2001-03-22	
; PRIOR APPLICATION NUMBER: 60/279,840	
; PRIOR FILING DATE: 2001-03-29	
; PRIOR APPLICATION NUMBER: 60/282,981	
; PRIOR FILING DATE: 2001-04-11	
; PRIOR APPLICATION NUMBER: 60/283,656	
; PRIOR FILING DATE: 2001-04-13	
; PRIOR APPLICATION NUMBER: 60/309,247	
; PRIOR FILING DATE: 2001-07-31	
; PRIOR APPLICATION NUMBER: 60/311,754	
; PRIOR FILING DATE: 2001-08-17	
; PRIOR APPLICATION NUMBER: 60/313,331	
; PRIOR FILING DATE: 2001-08-17	
; NUMBER OF SEQ ID NOS: 260	
; SOFTWARE: PatentIn Ver. 2.1	
; SEQ ID NO 42	
; LENGTH: 914	
; TYPE: PRT	
; ORGANISM: Homo sapiens	
US-10-032-189-42	
Query Match 20.6%; Score 952.5; DB 12; Length 914;	
Best Local Similarity 28.8%; Pred. No. 4.6e-82;	
Matches 260; Conservative 161; Mismatches 308; Indels 173; Gaps 31;	
Qy 74	RLVRITDADHPRWEVDIIPAPAGDVLDHAPPASSAPLQGRVLSPPAGSDLVLTVHAS 133
Db 83	RLKINEETPLKPRFEVDVLTSPSTVRLI-----SCSGDTGSLIADGKDLKCHITAN 137





491 TRDGSDEGCMWPGSAGYPDFNTPTMEAWNMESYDNYEGSAPNLFVNDNMKPSVFN 550  
 447 PEPNMLDDPPYRINNDGTGRPINNKTVRPLAVHYGGVTEYEEHNLFGLLEAATGSGVL 506  
 551 PE-VTMLKD-----AQHYGWEHRDVHNIYGLVHMTADGLR 587  
 507 RDTG--RRPFVLSSTFVSGRYTAVTGDNAATWGLDRLYSINTMLSFGLFGMPMIGADI 564  
 588 QRSQGMERPFVLAARFAGSQRFGAVMTGONTAEWDHLKISIPMCLSLGLVGLSFCGADV 647  
 565 CGFNGNTTEELCGRWIQLGAFYFESRDHSAIFTVRRELYLWPSVAAS-GRKALGLRYQLL 623  
 648 GGFPEKPEPELLVWYQMGAYQPFRAHAHLDTGRREPWLLPSQHNDIIRDALGQRYSL 707  
 624 PYFTLMEAHMTGAPITARPLFFSYPHDVATYGVDRQFLLRGVLSVPLVEPGFTTVDAY 683  
 708 PFWYTLTYQAHREGIPVWRPLWQYPODVTTFNIDDOYLLGDALLVHPVSDSGAHGVQVY 767  
 684 FPA-GR-WYRLDYSLAVATRTGKH-----VRLPAPADTVNVHLTGTTILPLQOQSALTTS 736  
 768 LPGGGEWYDIQSYQ-----KHGPGQTLVPLVTLSIPVFORGGTIVP-----RWM 813  
 737 RARRTA-----FHLVALAEDGTASGYLFLDDGDSPEYGRSDMSWVRFNKYIPNNKG 789  
 814 RVRSSCKMDPPTLFLVALSPQGTAAQGLFLDDGYTFNTQTRQEFLLRRFSF-----SG 868  
 790 AIKVSSEVHNSVAQSTLVISKVLMGHRSPAA 823  
 869 NTLVSSADPEGHFET-PIWIERVVIIGAGKPA 901

RESULT 12

US-10-032-189-122  
 ; Sequence 122, Application US/10032189  
 ; Publication No. US20030170630A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Alsobrook II, John P  
 ; APPLICANT: Tchernev, Velizar T  
 ; APPLICANT: Liu, Xiaohong  
 ; APPLICANT: Spytek, Kimberly A  
 ; APPLICANT: Zerrhusen, Bryan D  
 ; APPLICANT: Patturajan, Meera  
 ; APPLICANT: Grosse, William M  
 ; APPLICANT: Lepley, Denise M  
 ; APPLICANT: Burgess, Catherine E  
 ; APPLICANT: Shimkets, Richard A  
 ; APPLICANT: Grosse, William M  
 ; APPLICANT: Szekeres, Edward S  
 ; APPLICANT: Vernet, Corine A.M.  
 ; APPLICANT: Li, Li  
 ; APPLICANT: Casman, Stacie J  
 ; APPLICANT: Boldog, Ferenc L  
 ; APPLICANT: Gorman, Linda  
 ; APPLICANT: Gangolli, Esht A  
 ; APPLICANT: Fernandes, Elma R  
 ; APPLICANT: Rieger, Daniel K  
 ; APPLICANT: Edinger, Shlomit R  
 ; APPLICANT: Gunther, Erik  
 ; APPLICANT: Millet, Isabelle  
 ; APPLICANT: Sciore, Paul  
 ; APPLICANT: Ellerman, Karen  
 ; APPLICANT: MacDougall, John R  
 ; APPLICANT: Smithson, Glenda  
 ; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
 ; FILE REFERENCE: 21402-228  
 ; CURRENT APPLICATION NUMBER: US/10/032,189  
 ; CURRENT FILING DATE: 2001-12-21  
 ; PRIOR APPLICATION NUMBER: 60/257,495  
 ; PRIOR FILING DATE: 2000-12-21  
 ; PRIOR APPLICATION NUMBER: 60/258,171  
 ; PRIOR FILING DATE: 2000-12-20  
 ; PRIOR APPLICATION NUMBER: 60/269,940  
 ; PRIOR FILING DATE: 2001-02-20

; PRIOR APPLICATION NUMBER: 60/274,192  
 ; PRIOR FILING DATE: 2001-03-08  
 ; PRIOR APPLICATION NUMBER: 60/277,826  
 ; PRIOR FILING DATE: 2001-03-22  
 ; PRIOR APPLICATION NUMBER: 60/279,840  
 ; PRIOR FILING DATE: 2001-03-29  
 ; PRIOR APPLICATION NUMBER: 60/282,981  
 ; PRIOR FILING DATE: 2001-04-11  
 ; PRIOR APPLICATION NUMBER: 60/283,656  
 ; PRIOR FILING DATE: 2001-04-13  
 ; PRIOR APPLICATION NUMBER: 60/309,247  
 ; PRIOR FILING DATE: 2001-07-31  
 ; PRIOR APPLICATION NUMBER: 60/311,754  
 ; PRIOR FILING DATE: 2001-08-17  
 ; PRIOR APPLICATION NUMBER: 60/313,331  
 ; PRIOR FILING DATE: 2001-08-17  
 ; NUMBER OF SEQ ID NOS: 260  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 122  
 ; LENGTH: 944  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-032-189-122  
 Query Match 19.7%; Score 911.5; DB 12; Length 944;  
 Best Local Similarity 31.0%; Pred. No. 4.3e-78;  
 Matches 271; Conservative 114; Mismatches 296; Indels 193; Gaps 31;  
 QY 73 SRLRVRIITDADHPREVPQDIIPAPAGDVLHDAPPASSAPLQGR-----VLSPAGSDLV 127  
 DB 99 TRFRIDELEPRPRYRVP-----DVLVADPPARLSVSGRDENSVELTWAEGPYK 148  
 QY 128 LTVHASPPR-----PTVSRR-----STGD----- 146  
 DB 149 IILTARPPFLDLEDRLSLLSVNARGLLEFEBHQAPRVSGSKDPAEGDGAQPEETPRDG 208  
 QY 147 -----TLFDTP-----GLVFRDKYLEVTSALPACRASLYG 177  
 DB 209 DKPBETQKAEDEPGAWEETFKTHSDSKPGPMSVGLDF-----SLP-GMEHYG 258  
 QY 178 LGHEKTSFRLR--HNSDFTLWNADI--GASYVDVNLGYSHHPYMDVRAPGTAGHVL 232  
 DB 259 IPEHA-DNRLKVTGEGGPYRLNLDVQYELYNPMALYGSVPVLL-AHNPHRLDGLIFWL 316  
 QY 233 -----SSNG-----MDVLYG-----SYTYKVIQGVLDYFPAGNPLAVV 269  
 DB 317 NAAETWVDISNTAGKTLFGQMDYLGQSGETPOTDVRMMSSETGIIDVFLLLGSPISDVF 376  
 QY 270 DQYTLIARPAPEMPYWSFGFHQCRYGYNVSDLRVVRVAKARIPLEVMWTDIDYMDGF 329  
 DB 377 QYASLTGTQALPPLFSLYGHQSRWYRDEADVLEVDQGFDDHNLPCDVIWLDIEHADGK 436  
 QY 330 KDFLDRVNFPAELRPPVDRLHRNAOKYVILDPGRVDPIDATYGTFRGMOODIFLK 389  
 DB 437 RYFTWDPSPRP--PQPRMTLERLASKRKLVAIVDPHIKV---DSGYRVEELRLNGLHYVK 491  
 QY 390 -RNGTNFVGNVPGDVTFPDPMPHAAAEFAWRELSL--FRRTIPVDCGLWDMNEISFYN 446  
 DB 492 TRDGSDEGCMWPGSAGYPDFNTPTMEAWNMESYDNYEGSAPNLFVNDNMKPSVFN 551  
 QY 447 PEPNMLDDPPYRINNDGTGRPINNKTVRPLAVHYGGVTEYEEHNLFGLLEAATGSGVL 506  
 DB 552 PE-VTMLKD-----AQHYGWEHRDVHNIYGLVHMTADGLR 588  
 QY 507 RDTG--RRPFVLSSTFVSGRYTAVTGDNAATWGLDRLYSINTMLSFGLFGMPMIGADI 564  
 DB 589 QRSQGMERPFVLAARFAGSQRFGAVMTGONTAEWDHLKISIPMCLSLGLVGLSFCGADV 648  
 QY 565 CGFNGNTTEELCGRWIQLGAFYFESRDHSAIFTVRRELYLWPSVAAS-GRKALGLRYQLL 623  
 DB 649 GGFPEKPEPELLVWYQMGAYQPFRAHAHLDTGRREPWLLPSQHNDIIRDALGQRYSL 708  
 QY 624 PYFTLMEAHMTGAPITARPLFFSYPHDVATYGVDRQFLLRGVLSVPLVEPGFTTVDAY 683

Db 709 PFWYLLYQAHREGIPWMLRPLWYQDVTTFNIDDOYLLGDALLVHPVSDSGAHCVQVY 768  
Qy 684 FPA-GR-WRLXYDYLAVATRTGKH-----VRLPAPADTVNHLTGGTILPLQOQSALTTTS 736  
Db 769 LPOGGEVWYDIQSYQ-----KHHGPOTLYLPVTLSSIPVFORGGTIVP-----RMM 814  
Qy 737 RABRTA-----FHLVALAEDGTASGYLFLDDGDSPEYGRSDMSVMRENYKIPNNKG 789  
Db 815 RVRSECMKDDPITLFLVALSPQGTAGELFLDDGYTFNVTQKEFLLRFSF-----SG 869  
Qy 790 AIKVEVHNSYAQSRTLVISKVLMGHRSPAA 823  
Db 870 NTLVSSADPEGHET-PIWIERVVIIGACKPAA 902

RESULT 13

US-10-032-189-121  
; Sequence 121, Application US/10032189  
; Publication No. US20030170630A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsobrook II, John P  
; APPLICANT: Tchernev, Velizar T  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Zerhusen, Bryan D  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Grosse, William M  
; APPLICANT: Lepley, Denise M  
; APPLICANT: Burgess, Catharine E  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Grosse, William M  
; APPLICANT: Szekeres, Edward S  
; APPLICANT: Vernet, Corine A.M.  
; APPLICANT: Li, Li  
; APPLICANT: Casman, Stacie J  
; APPLICANT: Boldog, Ferenc L  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gangolli, Esha A  
; APPLICANT: Fernandes, Elma R  
; APPLICANT: Rieger, Daniel K  
; APPLICANT: Edinger, Shlomit R  
; APPLICANT: Gunther, Erik  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Sciore, Paul  
; APPLICANT: Ellerman, Karen  
; APPLICANT: Macdougall, John R  
; APPLICANT: Smithson, Glenda  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-228  
; CURRENT APPLICATION NUMBER: US/10/032,189  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 60/257,495  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/258,171  
; PRIOR FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: 60/269,940  
; PRIOR FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 60/274,192  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/277,826  
; PRIOR FILING DATE: 2001-03-22  
; PRIOR APPLICATION NUMBER: 60/279,840  
; PRIOR FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/282,981  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: 60/283,656  
; PRIOR FILING DATE: 2001-04-13  
; PRIOR APPLICATION NUMBER: 60/309,247  
; PRIOR FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: 60/311,754  
; PRIOR FILING DATE: 2001-08-17  
; PRIOR APPLICATION NUMBER: 60/313,331

; PRIOR FILING DATE: 2001-08-17  
; NUMBER OF SEQ ID NOS: 260  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 121  
; LENGTH: 966  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-10-032-189-121  
  
Query Match 19.5%; Score 903.5; DB 12; Length 966;  
Best Local Similarity 29.5%; Pred. No. 2.7e-77;  
Matches 265; Conservative 123; Mismatches 292; Indels 217; Gaps 31;  
  
Qy 73 SRLVRITDADHPREVPQDIIIPRAPGDVILHDAPPASSAPLQGR-----VLSPAGSDLV 127  
Db 99 TRIIRIDELEPRPRIRVP-----DVLVADPTARLSVSGRDDNSVELTVAEPYK 148  
  
Qy 128 LTVHASPPFR-----FTVSRRTGDTLFD--TAPGLVFRDKYLEVTSAL----- 168  
Db 149 IILTAQPPRLDLEDRLSLLSVNARGLMAFEHQAPRPVPSDK--VSLALGSVMDKIKN 205  
  
Qy 169 -----PA----- 170  
Db 206 LFSRQESKDPASGNGAQPPEATPGDGDKPEETQEKAEXDEPGAMEETFKTHSDSKPYOPTS 265  
  
Qy 171 -----GRASLYGLGHEHTKSSPRLR---HNDSTLWNADIGASYVD--VNLXGS---- 213  
Db 266 VGLDPSLPGEHVYGIPEHA-DSLRLKVTGEGEPRLYNLDVFOYBLNNPMALYGSVPVL 324  
  
Qy 214 --HPFYMDVRAPGTAHGVLL-----SSNG-----MDVLYGG-----SVYTY 248  
Db 325 LAHSFHRDL-----GIFMLNAAETWVDISSNTAGTKLFGKMLDYLQSGSETPQTOIRW 377  
  
Qy 249 KVIQGVLDFFEAGPNPLAVVDQYQTLLIARAPMPYSPFGHOCRYGVLYNSDLERVVAR 308  
Db 378 MSEGIIIDVFLMLGSPVDFRQYASLTGTALPLFLSLGYHQSRMWRDADVLEVDQG 437  
  
Qy 309 YAKARIPLVNMWTDIDYMDGFKPTLDRVNFNTAAELRP--FVDRLHRNAOKYVILDPGI 366  
Db 438 FDDHNMPCDVIWLDIEHADGKRYFTWDTFRP---QPLNMLEHLASKRKLVAIVDPI 493  
  
Qy 367 RVPIDATYGTFRGMQODIFLK-RNGTNFVGNVMPGDVYFPDFPMFPAAAEFWAREISL- 424  
Db 494 KV---DSCYRVHEELRNHGLYVKTRDSDGYEWCWPGSASYPDFTNPRMRAWNSMFSFD 550  
  
Qy 425 -PRRTIPVDGLWIDMNEISNFWNPEPMNALDPPYRINNDGTGRPINNKTYRPLAVHYGG 483  
Db 551 NYEGSAPNLYVNDMNEPSVFNGPE-VTMLKD-----AVHYGG 587  
  
Qy 484 VTEYEEHNLFLGLEARATGRGLRDTG--RRPFVLSRSTFVSGRYTAYWTGDNAATMGD 541  
Db 588 WEHRDIHNIYGLYVHMATAOGLIORSGGIERPFVLSRAFFSGSORFGAVMTGDNATAEDWH 647  
  
Qy 542 LKYSINTMLSFGLFPMGIMIGADI CGFNQNTTEELCGRWIQLGAPYFSPSRDHSALFTVRR 601  
Db 648 LKISIPMCLSLALYGLSFCGADVGGFKNPEPELLVRMYQMGAYQPFRAHAHLDTGRRE 707  
  
Qy 602 LYLWPSVAASG-RKALGLRYOLLPVFTYLMYEAHMTGAPIARPLFFSYPHDVATYGVDRQ 660  
Db 708 PMLLASQYQDAIRDALFORYSLLPFWYTLFYQAHKEGFPVNRPLWVQVPEDMSTFSEDQ 767  
  
Qy 661 FLGRGLVSPVLEFGPTTVTDAYFPAGR--MYRLDYSLAVATRTGKH-----VRLPAPA 713  
Db 768 PMLGDALLIHPVDAGAHGVQVYLPQGEEVWYDIQSYQ-----KHHGQTLVLPVTL 819  
  
Qy 714 DTNVHLTGGTILPLQOQSALTTSRARTA-----FHLVALAEDGTASGYLFLDDGDS 766  
Db 820 SSIPVFORGGTIVP-----RMMRVRRSSDCMKDDPITLFLVALSPQGTAGELFLDDGHT 873  
  
Qy 767 PEYGRSDMSVMRENYKIPNNKGAIKYKSEVHNSYAQSRTLVISKVLMGHRSPAA 823  
Db 874 FNYQTRHEFLLRFSF-----SGSTLVSSADPKGHLET-PIWIERVVIIGACKPAA 924





APPLICANT: Patturajan, Meera  
APPLICANT: Grosse, William M  
APPLICANT: Lepley, Denise M  
APPLICANT: Burgess, Catherine E  
APPLICANT: Shimkets, Richard A  
APPLICANT: Grosse, William M  
APPLICANT: Szekeres, Edward S  
APPLICANT: Vernet, Corine A.M.  
APPLICANT: Li, Li  
APPLICANT: Casman, Stacie J  
APPLICANT: Boldog, Ferenc L  
APPLICANT: Gorman, Linda  
APPLICANT: Gangolli, Bsha A  
APPLICANT: Fernandes, Elma R  
APPLICANT: Rieger, Daniel K  
APPLICANT: Edinger, Shlomit R  
APPLICANT: Gunther, Erik  
APPLICANT: Millet, Isabelle  
APPLICANT: Sciore, Paul  
APPLICANT: Ellerman, Karen  
APPLICANT: MacDougall, John R  
APPLICANT: Smithson, Glennda  
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-228  
CURRENT APPLICATION NUMBER: US/10/032,189  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: 60/257,495  
PRIOR FILING DATE: 2000-12-21  
PRIOR APPLICATION NUMBER: 60/258,171  
PRIOR FILING DATE: 2000-12-20  
PRIOR APPLICATION NUMBER: 60/269,940  
PRIOR FILING DATE: 2001-02-20  
PRIOR APPLICATION NUMBER: 60/274,192  
PRIOR FILING DATE: 2001-03-08  
PRIOR APPLICATION NUMBER: 60/277,826  
PRIOR FILING DATE: 2001-03-22  
PRIOR APPLICATION NUMBER: 60/279,840  
PRIOR FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: 60/282,981  
PRIOR FILING DATE: 2001-04-11  
PRIOR APPLICATION NUMBER: 60/283,656  
PRIOR FILING DATE: 2001-04-13  
PRIOR APPLICATION NUMBER: 60/309,247  
PRIOR FILING DATE: 2001-07-31  
PRIOR APPLICATION NUMBER: 60/311,754  
PRIOR FILING DATE: 2001-08-17  
PRIOR APPLICATION NUMBER: 60/313,331  
PRIOR FILING DATE: 2001-08-17  
NUMBER OF SEQ ID NOS: 260  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 124  
LENGTH: 944  
TYPE: PRT  
ORGANISM: Sus scrofa  
US-10-032-189-124

Query Match 19.0%; Score 878.5; DB 12; Length 944;  
Best Local Similarity 33.0%; Pred.No. 6.5e-75;  
Matches 256; Conservative 104; Mismatches 294; Indels 121; Gaps 31;  
Qy 95 PRPAGDVLHDAPPASSAPLQGRVLSPAGSDLVLTVHASPRFTSVRRSTGDTLFDTPAG 154  
Db 201 PEEAPGD-GDKP-----EEIQGK-----ABKD-----EPGAWETFTKTHSDSKPYGPTSVG 245  
Qy 155 LVFRDKLYVTSALPAGRASLYLGCHTKSFLRL-----HNDSFTLWNADI--GASYVDVN 209  
Db 246 LDF-----SLP-GMEHYGIPEHA-DSLRLKVTGGDPYRLNLDVYQYELYNPMA 294  
Qy 210 LYGSHPFYMDVRAGTAHGVLL-----SSNG-----MDVLYGG-----SYV 246  
Db 295 LYGSVPVLL-AHSPHRLDGLFNLNAEATWVDISSNTAGKTLFGKMLDYLQSGGETPTQDV 353  
Qy 247 TYKVIIGVLDYFFAGPNPLAVVDQYTQLIARPAPMPYSGFGHQCRGYGLNVSDLERVV 306

Db 354 RMSESGLIDVFLLLGSPSPDVFQYASLTCTQALPPLFSLYGYSRWNRYDEADVLEVN 413  
Qy 307 ARYAKARIPLEVMMTDIDYMGKDFDLDLVNFTAAELRPFVDRHLRHAQKYLILDPGI 366  
Db 414 QGFDDHNLPCDFIWLDIRHADGKRYFTWDPGRF--POPRTMLEHLASKRKLVAIVDPHI 471  
Qy 367 RVDPIDATYGTFRGMQODIFLK-RNGTNNFVGNVWPGDVYFPDFMHPAAAFWAREISLF 425  
Db 472 KY---DSSYRVHEELQNLGLYVKTRDSDYEGWCWPGGAASYPDFTNPKORAWNA---DMF 525  
Qy 426 RRTIPVDG-----LWIDMNEISNFPNPEPMNALDDPPYRINNDGTGRPINNKTVRPLAV 479  
Db 526 RFE-NYEGSSSNLYVMNDMNEPSVFNGPE-VTMLKD-----AQ 561  
Qy 480 HYGVTYEENLFGLEARATGRG-VLRDTG-RRPVFLSRSTFVSGRYTAYWTGNAA 537  
Db 562 HYGWEHRDLHNIYGFYVHMATADGLVLRSGGVERPFVLSRAFFAGSQRFQAVMTGNTA 621  
Qy 538 TWGDLRYSINTMLSFGLFGMPMIGADICGFNGNTTEELCGRWIOLGAFYFCSRDHSAIFT 597  
Db 622 EWDHLKISIPMCLSLGLVGVSEFCGADVGGFFKNPEPELLVRYQMGAYQPFRAHAHLD 681  
Qy 598 VRRELYLWPSVAASG-RKALGLRYQLLPFYTYLMEAHMTGAPLARPLFFSYPHDVATYG 656  
Db 682 GRREPMLLPQYQDMIRDALGQYSLLPFWYTLFYQAHREGVPVNRALWVHYPODVTTFS 741  
Qy 657 VDRQFLGRGVLSVPVLEPGPTTVDAYPPA-GR-WYRLYDYSLAVATRTGKXVRLPAPAD 714  
Db 742 IDDEFLLGDALLVHFVTDSEAHGVQVYLPQGQEVVYDVHSYQKYHGQPT---LYLPVTL 798  
Qy 715 TVNVHLTGTTILPQQSALTTSRARPTA-----FHLVALAEDGTAGSYLFLDDGDS 767  
Db 799 SIPVFORGGTIVP-----RNRVRRSSDCMKDDPITLFLVALSPQTAQGBFLDDGHTF 852  
Qy 768 EYGRSDWSMVRFNKYIPNNKGAIKVKSEVHNYSYAQSRTLVISKVLMGHRSPA 822  
Db 853 NYQTGHEFLRRFSP-----SGNTLVSSSADSCKGHFET-PVMIRVVIIGAGKPA 901

Search completed: October 27, 2003, 10:38:07  
Job time : 110.357 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 ; Search time 15.8523 Seconds  
(without alignments)  
2340.767 Million cell updates/sec

Title: US-10-043-418-1  
Perfect score: 4630  
Sequence: 1 MATVGVLLCLCLCLFAPRL.....ICGLSLVVGEEELKVMASY 877

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4630	100.0	877	1 US-08-430-925A-4	Sequence 4, Appli
2	883	19.1	914	4 US-09-437-054A-8	Sequence 8, Appli
3	826.5	17.9	919	4 US-09-437-054A-17	Sequence 17, Appl
4	723.5	15.6	938	4 US-08-897-843A-1	Sequence 1, Appli
5	679.5	14.7	693	4 US-09-376-343-2	Sequence 2, Appli
6	466	10.1	1066	2 US-08-633-770A-1	Sequence 1, Appli
7	464.5	10.0	1070	2 US-08-633-770A-2	Sequence 2, Appli
8	385.5	8.3	1092	4 US-09-275-608-3	Sequence 3, Appli
9	356	7.7	1091	3 US-08-633-768A-2	Sequence 2, Appli
10	349	7.5	1088	3 US-08-633-768A-1	Sequence 1, Appli
11	279	6.0	570	4 US-09-275-608-4	Sequence 4, Appli
12	197	4.3	390	4 US-09-634-238-331	Sequence 331, App
13	195.5	4.2	150	4 US-09-437-054A-10	Sequence 10, Appli
14	132.5	2.6	6396	4 US-09-410-551B-72	Sequence 72, Appl
15	121.5	2.6	692	4 US-09-198-452A-966	Sequence 966, App
16	121	2.6	3729	2 US-08-804-227C-4	Sequence 4, Appli
17	119	2.6	613	4 US-09-149-727-5	Sequence 5, Appli
18	119	2.6	651	4 US-09-715-858-2	Sequence 2, Appli
19	116	2.5	3816	3 US-09-438-517-3	Sequence 3, Appli
20	113.5	2.5	286	4 US-09-634-238-330	Sequence 330, App
21	111	2.4	1578	4 US-09-410-551B-31	Sequence 31, Appl
22	111	2.4	1706	4 US-08-669-785-2	Sequence 2, Appli
23	110.5	2.4	3201	4 US-09-679-279-15	Sequence 15, Appl
24	109.5	2.4	1711	2 US-08-342-930-2	Sequence 2, Appli
25	109	2.4	979	4 US-08-514-213A-2	Sequence 2, Appli
26	108	2.3	1794	6 5183745-6	Patent No. 5183745
27	107	2.3	1482	4 US-09-410-551B-21	Sequence 21, Appl

28	107	2.3	1489	6 5183745-2	Patent No. 5183745
29	105.5	2.3	814	4 US-09-328-352-4373	Sequence 4373, Ap
30	104.5	2.3	2930	4 US-09-417-822-2	Sequence 2, Appli
31	104	2.2	433	4 US-09-104-623A-4	Sequence 4, Appli
32	104	2.2	433	4 US-09-019-532-4	Sequence 4, Appli
33	104	2.2	635	2 US-08-873-479-43	Sequence 43, Appli
34	104	2.2	883	4 US-09-252-991A-16929	Sequence 16929, A
35	104	2.2	3562	4 US-09-679-279-14	Sequence 14, Appli
36	103.5	2.2	1705	4 US-08-669-785-4	Sequence 4, Appli
37	101.5	2.2	1848	3 US-08-296-791-6	Sequence 6, Appli
38	101.5	2.2	1848	5 PCT-US95-10661A-6	Sequence 6, Appli
39	100.5	2.2	617	4 US-09-252-991A-22318	Sequence 22318, A
40	100.5	2.2	894	4 US-09-252-991A-21126	Sequence 21126, A
41	100.5	2.2	983	2 US-08-164-292B-26	Sequence 26, Appli
42	100.5	2.2	983	3 US-08-845-623-26	Sequence 26, Appli
43	100.5	2.2	983	3 US-08-815-927-26	Sequence 26, Appli
44	100.5	2.2	983	4 US-09-103-330-26	Sequence 26, Appli
45	100.5	2.2	983	4 US-09-435-242-26	Sequence 26, Appli

ALIGNMENTS

RESULT 1  
US-08-430-925A-4  
; Sequence 4, Application US/08430925A  
; Patent No. 5763252  
; GENERAL INFORMATION:  
; APPLICANT: Skadsen, Ronald W  
; APPLICANT: Tibbot, Brian K  
; TITLE OF INVENTION: Cloned Alpha-Glucosidase from Barley  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: 1 South Pinckney Street  
; CITY: Madison  
; STATE: WI  
; COUNTRY: US  
; ZIP: 53703  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/430,925A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J  
; REGISTRATION NUMBER: 27,386  
; REFERENCE/DOCKET NUMBER: 960296.93171  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 608-251-5000  
; TELEFAX: 608-251-9166  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 877 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-430-925A-4

Query Match 100.0%; Score 4630; DB 1; Length 877;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 877; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MATVGVLLCLCLCLFAPRLCSKSGPLAARTVLA VVTMEGALRAEAATGCRSSTGIV 60  
DB 1 MATVGVLLCLCLCLFAPRLCSKSGPLAARTVLA VVTMEGALRAEAATGCRSSTGIV 60  
QY 61 QRLAVASLETDSRLVRITDADHPHWEVQDIIIPAPGVDVLDHAPPASSAPLQGRVLS 120

Db 61 QRLAVYASLETDLSRLVRITDADHPWEVPEQDII PRPAPGVLDHAPPASSAPLQGRVLS 120  
Qy 121 PAGSDLVLTWHASPPFTVSRRSTGDTLFTAPGLVFRDKYLETVSALPAGRASLYGLGE 180  
Db 121 PAGSDLVLTWHASPPFTVSRRSTGDTLFTAPGLVFRDKYLETVSALPAGRASLYGLGE 180  
Qy 181 HTKSSFLRHNDSTFLWNADIGASYVDVNLVYGSHPFVMDVRAPGTAHGVLLSSNGMDVL 240  
Db 181 HTKSSFLRHNDSTFLWNADIGASYVDVNLVYGSHPFVMDVRAPGTAHGVLLSSNGMDVL 240  
Qy 241 YGGSVYTVKIVGGVLDYFFAGNPLAVVDQYTLIARPAPEYVFGFHQCRGYLNVLS 300  
Db 241 YGGSVYTVKIVGGVLDYFFAGNPLAVVDQYTLIARPAPEYVFGFHQCRGYLNVLS 300  
Qy 301 DLERVARYAKARIPLEVMKTDIDYMDGPKDFTLDRVNFETAAELRPFVDRLHRNAOKYVL 360  
Db 301 DLERVARYAKARIPLEVMKTDIDYMDGPKDFTLDRVNFETAAELRPFVDRLHRNAOKYVL 360  
Qy 361 ILDPGIRVDPIDATYGTFRVGMQODIFLKRNGTNFVGNWPGDVYFPDFMHPAAAEFWAR 420  
Db 361 ILDPGIRVDPIDATYGTFRVGMQODIFLKRNGTNFVGNWPGDVYFPDFMHPAAAEFWAR 420  
Qy 421 EISLFRRTIPVDGLWIDMNEISNFYNPEPMNALDDPPYRINNDGTGRPINNKTVRPLAVH 480  
Db 421 EISLFRRTIPVDGLWIDMNEISNFYNPEPMNALDDPPYRINNDGTGRPINNKTVRPLAVH 480  
Qy 481 YGGVTEYEENHFLGLEARATGRVLRDTCRRPVLRSFTVSGSRYTAYMTGDNAAWTG 540  
Db 481 YGGVTEYEENHFLGLEARATGRVLRDTCRRPVLRSFTVSGSRYTAYMTGDNAAWTG 540  
Qy 541 DLRYISINTMLSFGLFGPMIGADICGPNNGTTBELCGRWIQLGAFYPPFGRDHSIAIFTVR 600  
Db 541 DLRYISINTMLSFGLFGPMIGADICGPNNGTTBELCGRWIQLGAFYPPFGRDHSIAIFTVR 600  
Qy 601 ELVLPWSAASGRKALGRYQLLPYFTLMEYAHMTGAPIARPLFFSPHVDATYGVDRQ 660  
Db 601 ELVLPWSAASGRKALGRYQLLPYFTLMEYAHMTGAPIARPLFFSPHVDATYGVDRQ 660  
Qy 661 FLGRGVLSVPLEPGPTTVDAYFPAGRWRLDYSLAVATRGKHVLPAPADTVNVHL 720  
Db 661 FLGRGVLSVPLEPGPTTVDAYFPAGRWRLDYSLAVATRGKHVLPAPADTVNVHL 720  
Qy 721 TGGTILPQOSALTTSRRRTAFHLLVALAEDGTASGYLFLDDGDSPEYGRSDMSVRF 780  
Db 721 TGGTILPQOSALTTSRRRTAFHLLVALAEDGTASGYLFLDDGDSPEYGRSDMSVRF 780  
Qy 781 NYKIPNNKGAIKVSEVYVHNSYAGSRTLVLISKVLMGHRSPAAPKKLTVHNSAEYAS 840  
Db 781 NYKIPNNKGAIKVSEVYVHNSYAGSRTLVLISKVLMGHRSPAAPKKLTVHNSAEYAS 840  
Qy 841 SAGTRYONAGLGGVAHIGGLSLVVGEEFELKVMSY 877  
Db 841 SAGTRYONAGLGGVAHIGGLSLVVGEEFELKVMSY 877

## RESULT 2

US-09-437-054A-8  
; Sequence 8, Application US/09437054A  
; Patent No. 6316698  
; GENERAL INFORMATION:  
; APPLICANT: Allen, Stephen M.  
; APPLICANT: Kinney, Anthony J.  
; TITLE OF INVENTION: Plant Alpha-Glucosidase II Homologs  
; FILE REFERENCE: B81273 US NA  
; CURRENT APPLICATION NUMBER: US/09/437,054A  
; CURRENT FILING DATE: 2001-05-14  
; PRIOR APPLICATION NUMBER: 60/107,909  
; PRIOR FILING DATE: 1998-No. 6316698ember-10  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 8  
; LENGTH: 914  
; TYPE: PRT

; ORGANISM: Glycine max  
US-09-437-054A-8  
Query Match 19.1%; Score 883; DB 4; Length 914;  
Best Local Similarity 29.4%; Pred. No. 6.8e-78;  
Matches 278; Conservative 123; Mismatches 338; Indels 206; Gaps 36;  
Qy 7 LLLCLCLCLFAPRLCSSKEE-----GPLAAR-----TVLAVAVTM-EGALPABA 49  
Db 8 LLLLLLCLSHLSVLSWKKEEFRTCHQTPFKRARSAPGSSSLIATDVTISHGDLTAK- 66  
Qy 50 ATGCRSSTGDVORLAVYASLETDLSRLVRITDAD-----HPRMEVQDII PRPAPCEVL 103  
Db 67 LTPKHDQSETKPILLTLVSVQRGILRLKI-DEPSSLSPPKRFEVP-DIVSEFPSTKL 124  
Qy 104 HDAPPASAP--LQGRVLSPAGSDLVLTWHASPPFTVSRRSTGOD-----TLFD--- 150  
Db 125 W-LPKISSVENGLSSVYVLSDGHSAVL--RHDPPPELFIRODSSGDRVISLNSHOLFDFEQ 181  
Qy 151 ---TAPGLVFRDKYLEVTSALPAGRASI-----YGLGHTKSSPRLR----- 189  
Db 182 LKHSEDDNWEEOQRSHTRDRPYGPQISFOVSFYGADFVYGPIDERA-ASLALKPTGPN 240  
Qy 190 --HNSDFTLNADICASYV---DYNLYGSHPP-----YMDYRAP 223  
Db 241 VDESEPYRLFNLDV-FEYIHDSPPFLYGSIPFMVSHGKARSSGFFFWLNAAEQMIDVLAP 299  
Qy 224 G-TAHGVULLLSNGMDVLYGGSYVTVYKIVGGVLDYFFAGNPLAVVDQYTLIARPAHM 282  
Db 300 GWDAESGIALPSHRIDTFPMSE-----AGVDPAFFIIGNPKVDVLRQYTAVTGTTPAMP 352  
Qy 283 PYWFGPHQCRGYGLVNSDLEVRVARYAKARIPLVNMWTDIDYMDGFKDFTLDRVNTAA 342  
Db 353 QLFSIAHYQCRWNTYRDEEDVEHVDSEKFDLIDPVDVLWLDIEHTDGKRYFTWDRALPHP 412  
Qy 343 ELRPFVDRLHRNAOKYVLLDPGIRVDPIDATYGTFRVGMQODIFLK-RNGTNFVGNWVP 401  
Db 413 E--EMQRKLASKGRHMVTIIVDPHIK--RDENFHLHKEASQKGVYVKDASGNDPDGWCW 467  
Qy 402 GDVFPDPMPHAAAFAFWARELSL--FRRTIPVDGLWIDMNEISNFYNPEPMNALDDPPYR 459  
Db 468 GSSYPDTLNPFEIRSMWADKFSYQSEGSTFSLYIWNDMNPSVFNPEVMTMPRD----- 522  
Qy 460 INNDGTGRPINNKTVRPLAVHYGGVTRYEENHFLGLEARATGRGVLR--DTGRRPFVLS 517  
Db 523 -----VTHYGGVEHRELHNAKYVYFHMATANGLLKRGESNDRPFVLS 564  
Qy 518 RSTFVSGSRYTAYTGNAAATWGDLYSINTMLSFGLFGPMIGADICGPNNGTTBELCG 577  
Db 565 RALFAGSORYGAVMTGONTADWDHLRVSIPMVLITGLTGMSPSGADICGFFGNPEPELLV 624  
Qy 578 RWIQLGAFYPPFSRDHSAIFTVRRELVLWPSVAASGRK-----ALGLRYOLLVPYVTL 630  
Db 625 RWYQLGAYYPPFRAHAHHTKRRPWLFF-----GENTELIKDAIHVRVALLPYFTLTP 678  
Qy 631 YEAMHTGAPIARPLFFSPHVDATYGVDRQFLGRGVLSVPLEPGPTTVDAYPAGRW 689  
Db 679 REANTTGVVVVRPLWMEFPSPDEATFSDNETFMWGSILVQGIYTERAKHASVYLPQKQSW 738  
Qy 690 YRLDYSLAVATRGKHVLPAPADTVNVHLTGTILPQOSALTTSRRRTAFHLLVALAED 742  
Db 739 ---YDLRTGAVYKGVVTHKLEVTESIPAFORAGTIIARK-----DRFRSSSTOMANDP 789  
Qy 743 FHLVALAEDGTASGYLFLDDGDSPEYGRSDMSVRFNYKIPNNKGAIKVSEVYHNSY 802  
Db 790 YTLVVALNSSOAEGLYIDDCSS-----FNF-----LQGYIHRFP 826  
Qy 803 AQSR-----TLWISKVYLMGHRSPAAPKKL 827  
Db 827 IFSNGKLTSLDAPASSSKGRYPSPDAFIERIILGH-APSSKNAL 870

RESULT 3

```
US-09-437-054A-17
; Sequence 17, Application US/09437054A
; Patent No. 6316698
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; TITLE OF INVENTION: Plant Alpha-Glucosidase II Homologs
; FILE REFERENCE: BB1273 US NA
; CURRENT APPLICATION NUMBER: US/09/437,054A
; PRIOR FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: 60/107,909
; PRIOR FILING DATE: 1998-No. 6316698ember-10
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 17
; LENGTH: 919
; TYPE: PRT
; ORGANISM: Solanum tuberosum
US-09-437-054A-17

Query Match      17.94; Score 826.5; DB 4; Length 919;
Best Local Similarity 28.04; Pred. No. 2.7e-72;
Matches 255; Conservative 136; Mismatches 330; Indels 189; Gaps 31;

QY 24 KEGPLAAR---TVLAVATVMEGALRAEAATGGRSSTGDVORLAVAYASLETDSRLVRKI 79
Db 69 KEENPESEQNKPLVLTSLVYQGVVRVK-----IDEDQNL----- 104
QY 80 TDADHPRWVPODI-----IPRPAGDVLLDAPPA 109
Db 105 -NPPKRFPEVTEEDFNLTKLWLRVKEBQIDGVSSPSSFYLSGDYEGVLRHD----- 159
QY 110 SSAPLOGRVLSPAGSDVLVTHASP--FRFTVSR--RSTGDTLFDTPAGLVRDRKYLEVTS 167
Db 160 ---PFEVFARESGKRVLSINGLDFEQLREKKGDD-----WEKFRSHDT 207
QY 168 LPAGRASL-----YLGHTKSSFRUL-----HNDSFTLNNADIGASTV 206
Db 208 RPYGPQISFDVSFYGADFVYGIPEHA--TSFALKPTKGPVVEEYSEPYRLFNLDV--FEYL 265
QY 207 ---DVNLIGHPPFMDVRAGTAGHVLLSSNGM--DVLVYGG--SVYTYKVI----- 251
Db 266 HESPFGLYGSIPPMISHGKARGSGFFMLNAEAMQIDVLGSGWNSDESSKIMLPDSKHRI 325
QY 252 -----GGVLDFFYFAGNPFLAVDQYQTLIARPAFMPYWSFGFHQCRGYLVNVDLSR 304
Db 326 DTLWMSSEGVVDTFPIGPGKDVVRQYTSVTGEPSPQLFATAYHQCWNRYADEEDVTN 385
QY 305 VVARYAKARIPLEVNWTDIDYMOGFKDFTLDRVNFTHAELRPPVDRLHRNAQKVVILDP 364
Db 386 VDSKFEHDIDPYDLVLDIEHTDGGKFTWDRVLFNPE--EMQKLAAGRHMTVIDP 443
QY 365 GIRVDPIDATYGTFRGMQODIFLK--RNGTNFVGNVWPGDVYFPDFMHPAAAEFWAREIS 423
Db 444 HIK---RDESHIPEALEKGYVYKDATGDKYDGCWCPGSSSYTDLLNPEIKSWNSDKES 500
QY 424 L--FRPTIPDGLWIDWNEISNFNPEPMNALDPPYRINNDGTGRPINNKTVRPLAVHY 481
Db 501 LDSVVGSTKLYLIWNNDNEPSVFNGPEVTMPD-----ALHH 537
QY 482 GGVTEYEHLPLGLEARATGRVLR--DTGRRPFLVLSRSTFVSGRYTAYWTCNAAFW 539
Db 538 GGVEHRELHNSYGYFFHMGTSGLLKGKGDGKORPFVLARAFFAGSQRYGAIWTDNTAEW 597
QY 540 GDLRYISINTLSFGLFGMPMIGADICFNGNTTEELCGRWIQALCAFYFSDRHSALFTVR 599
Db 598 EHLRVSPVMTLTSLSGVTFSGADVGGFFGNPDTELLVRVYQVGYAYYFFRGHAHDTKR 657
QY 600 RELYIWPSSVAAS--GRKALGLRYQLLPYFTYLMYFAHMTGAPIARPPLFFSYPHDVATYVD 658
Db 658 REPWLFGERNQLMREAIHVRMYLPLPYFTLFRANSNGTVPARPLMWEPPGDEKFSND 717
QY 659 RQFLLRGCVLSVPLEPGPTTVDAFFAGR--WYRLYDYSLAVATRTGKHVRLPAPADTVN 717

US-08-897-843A-1
; Sequence 1, Application US/08897843A
; Patent No. 6514493
; GENERAL INFORMATION:
; APPLICANT: Deleo, Albert B.; Loftus, Douglas; Appella, Ettore
; TITLE OF INVENTION: CDNA CLONE FOR MURINE TUMOR
; TITLE OF INVENTION: REJECTION ANTIGEN GP110 AND TUMOR PEPTIDE VACCINE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Diane R. Meyere
; STREET: 600 Grant Street, 42nd Floor
; CITY: Pittsburgh
; STATE: PA
; COUNTRY: USA
; ZIP: 15219
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/897,843A
; FILING DATE:
; CLASSIFICATION: 424
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 938 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE:
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Mus musculus
; DEVELOPMENTAL STAGE: Embryo
; TISSUE TYPE: Embryo
; CELL LINE: NIH 3T3
US-08-897-843A-1

Query Match      15.64; Score 723.5; DB 4; Length 938;
Best Local Similarity 27.64; Pred. No. 4.4e-62;
Matches 243; Conservative 127; Mismatches 299; Indels 211; Gaps 35;

QY 76 RVRITDAD--HPRWEVPODIIPRPAGDVLDHAPASSAPLOGR-----VLSPAGSDLVLT 129
Db 96 RIRIDELEPRPRVRP-----DVLVADPPPTARLSVSGRDDNSVELTVAEGPYKII 145
QY 130 VHASPFR-----FTVSR-----STGDT 147
Db 146 LTAQPFRLDLEDSLLSVNAGLMAFEPQAPRVPQBSKDPAGCQAQPEATPGDGK 205
QY 148 LFDT-----APGLVPRDKYLEVTSALPAGRAS-----LYGLGHTKSSFRL 188
Db 206 PEETOEAEXDEPG--AMEETFKTHSDSKPYGPTSVGLDFSLPGMEHVYGIPEHA--DSLRL 263
```



COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/633,770A  
FILING DATE: July 8, 1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP94/03398  
FILING DATE: OCT-15-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Altman, Daniel E  
REGISTRATION NUMBER: 34,115  
REFERENCE/DOCKET NUMBER: DY06.001APC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 714-760-0404  
TELEFAX: 714-760-9502  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1066 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-633-770A-1

Query Match 10.1%; Score 466; DB 2; Length 1066;  
Best Local Similarity 23.4%; Pred. No. 1.7e-36;  
Matches 212; Conservative 119; Mismatches 343; Indels 232; Gaps 39;

QY 100 GDLVLDAPPASPLQGRVLSPAGSDLVLTVAHSPRFTVSR-----141  
DB 119 GDFTFSSKTVAEKSETRNKVGDGRILHLWSPPFIQVVRILTLPKDPYIPNVAAAE 178  
QY 142 -RSTGDLFTAPGL-----VPRDKYLEVTSALPAGRASLYGLGHTKSSFLRLH 190  
DB 179 ARVSDKVMVOTSPKTFERKLNHPQHKMLKDTVLDI--VKPGHGEYVGVGEMGGIQF-MKE 234  
QY 191 NDSFTLNADIGASYVDV-----NLYGSHFFYMDVAPG-----TAHGVLLSSN 235  
DB 235 PTFNRYNFD-NMQOQVYVYAGALDSREPLYHSDPFYLDVNSNPENHKNITATIDNYSQI 293  
QY 236 GMDVLYGSGVYTVKIG---GVLDFFYFAGNPLAVVDQVTLIARPAAPMPYKSPGFHOC 292  
DB 294 AID--FGKTNISGYIKLTRYGGIDCYGISADTVPEIVRLYGLVGRSKLKPRIILGAHA 351  
QY 293 RYGLNVSDLERVVARAKARIPLEVNMWTDIDYMDGKDTLDRVNFATAELRPFVDRLLH 352  
DB 352 CYGYQESDLYSVVQYRDCCKFPLDGIHVDVDQDGFRTFTTNPHTF--PNPKEMFTNLR 409  
QY 353 RNAQKYVLIDPGRVDPIDATYGTFFVGMQODIFLK---RNGTNFVGN-----398  
DB 410 NNGIKCSTNITPVISINNREGCYSTLLEGVDKXKFIIMDDRYTEGTS--GNAKDVRYMYVG 467  
QY 399 -----VWPGDV-----YFPDFMHPAAA 415  
DB 468 GGNKVEVDNDVNGRDPDKNDYDFANFNSKQYPHGVSYGNGSAGFYDPLNRKEVR 527  
QY 416 EFWAREIS-LFRRTIPVDGLWIDWNEISNFNPEPMALDDPPYRINDGTGRPINNKTV 474  
DB 528 IWMGQVQKYLK--DMGLEFVWQMDMTTFAHTSYGDMKGL---PTRLI--VTSQVNTASE 580  
QY 475 RPLAVHYGGVTEYEHLNFGLEARATGRGVLRTDGR---RPFVLSRSTFVSGRVTAYW 531  
DB 581 KKLAIETWALYSYNLH-----KATWGLSRLESKKNKFNILSGSYAGAYRAGLW 632  
QY 532 TGDNAATWGLDLYRYSINTWLSGLFGPMIGADICGF-----NGNTE-----ELCGRIQ 581  
DB 633 TGDNASNWEFKISVSQVLSGLINGVCIAGSDTGGFEPYRDANGVEEKYCSPELLIRWT 692  
QY 582 LGAFYFSDHSAIFTVPR-----ELYLWPSVAASGRKALGL 618

DB 693 GSFLPLMLNRH---YVKDRKMFQBPYSYPKSHLETHPELADQAWLYKSVLEICRYVVEL 748  
QY 619 RYQLLPYFTLMEAHMTGAPIARPLFFSYPHDVATYG-----VDRQFLLRGVLVSPVL 673  
DB 749 RYSLIQLLYDCMFCQNVVDGMPITRSMLLTDTEDTTFENBSQKFLDNQYMGADDILVAPIL 808  
QY 674 E-----PGPTTVDAYFPA-GRNY-----RLYD-----YSLAVATRTGKV-----RLPAPAD--- 714  
DB 809 HSRKEIFGENR-DVYLPYHTWTPSNLRPMDQDQVALGNFVEGGSVINTARTVAPEADYN 867  
QY 715 ----TVAHLTGTTILPQOSALTTSR--ARRTAFLLVALAEDGTASGY-LFLDDG--- 764  
DB 868 LFHSVVPVYVREGAIIPQIEVRQWTCGGANRLKFNII-----PGDKKEYCTYLLDGVSR 922  
QY 765 -----DSPEYGRSRDSMW-----RFNYKI-----PNNKGAIKVKSEVVHNSYQ 804  
DB 923 DSAPEDLPQYKETHQSKVEGAETAKQIGKKTGYNISGTDPEAKGYHR-KVAVTQTSDK 981  
QY 805 SRTLVI 810  
DB 982 TRTVTI 987

RESULT 7  
US-08-633-770A-2  
; Sequence 2, Application US/08633770A  
; Patent No. 5908760  
; GENERAL INFORMATION:  
; APPLICANT: Bojeen, Kirsten  
; APPLICANT: Yu, Shukun  
; APPLICANT: Kragh, Karsten  
; APPLICANT: Christensen, Tove  
; APPLICANT: Marcussen, Jan  
; TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM A FUNGUS, ITS  
; TITLE OF INVENTION: PURIFICATION GENE CLONING AND EXPRESSION IN MICROORGANISMS  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson & Bear  
; STREET: 620 Newport Center Drive 16th Floor  
; CITY: Newport Beach  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/633,770A  
; FILING DATE: July 8, 1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP94/03398  
; FILING DATE: OCT-15-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Altman, Daniel E  
; REGISTRATION NUMBER: 34,115  
; REFERENCE/DOCKET NUMBER: DY06.001APC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 714-760-0404  
; TELEFAX: 714-760-9502  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1070 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-633-770A-2

Query Match 10.0%; Score 464.5; DB 2; Length 1070;



Db 577 DWGRPDVAQWNGENYKLF--SIGLDFVQDMTVPAMPHRLGDVANKN--SGSSAPGWP 632  
Qy 460 INNDGTGRPINNKVTP-----LAVHYGGVTEYEHNLFGLLEPARATGR-GVLRDTG 510  
Db 633 NENDPSNGRYNMKSYHPQVLVDMRYGAEGREPMVSQRNIHAYTLCESTRREGIVGNAD 692  
Qy 511 -----REPFLVLSSTFGSGRYTAYMTGDNAATMGDLRYSLNTMLSPGLFGWPMIGADIC 565  
Db 693 SLTKFRRSYIIISRGYIGNQHFQGMWGDNSATESYLOMMLANIINNNMSCPLVGSIDG 752  
Qy 566 GF-----NGNTTEELCGRMWQALQAFYFFGRDHSALFTVRR-----ELYLMPSSVAASG 612  
Db 753 GFTQYNDAGDPTEDLWRFVQAGCLLPWFNHYDRWIESKHKGYQELMYPGQKDTL 812  
Qy 613 RKALGLRYQLLPYFLMYEAHMTGAPI--ARLPFSYPHDVATY-GVDRQFL-----GR 665  
Db 813 KKFVEPRYQWELVLTAMYQNAITGEPHKAAPM--YNNVNVYKSONDHFLGHDGY 869  
Qy 666 GVLVSPVLEPGPTTVDAYFA-GRWYR-----LYDYSLAVATRTCK 705  
Db 870 RILCAPVREWATSREVLFPYKWFEGDPDTKPLENIEQGTLYNVA----- 920  
Qy 706 HVRLPAPADTVNHLTGTTLPLOQSALTTSRART-----AFHLLVALAEDGTASGY 758  
Db 921 -----APLNDSPFVREGTILPTRYTLTGYNKSNITVTDNDPLVFELEP--LENNQAGL 973  
Qy 759 LFLDDG----DSPEYGRSDWSMVRFNKYIPNNKGAIKVKEVHNSYQAQSRTLVISKV 814  
Db 974 FYHDGQVTTNAEDFGK--YSVISVKAAGEGSQMSVKFDFNEVHEQWASFTVRV---- 1026  
Qy 815 LMGRSPAAPKLTJV--HVNSAVEASS 840  
Db 1027 ---RNGAFSINNVSSQIGQDMQOQS 1050

## RESULT 9

US-08-633-768A-2  
Sequence 2, Application US/08633768A  
Patent No. 6013504  
GENERAL INFORMATION:  
APPLICANT: YU, SHUKUN  
APPLICANT: BOUSEN, KIRSTEN  
APPLICANT: KRAGH, KARSTEN  
APPLICANT: BOJKO, MAJA  
APPLICANT: NIELSEN, JOHN  
APPLICANT: MARGUSEN, JAN  
TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM  
TITLE OF INVENTION: A FUNGUS INFECTED ALGAE, ITS PURIFICATION  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/633.768A  
FILING DATE: 02-JUL-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 9321301.5  
FILING DATE: 15-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Altman, Daniel E  
REGISTRATION NUMBER: 34,115  
REFERENCE/DOCKET NUMBER: DY007.001APC  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 714-760-0404  
TELEFAX: 714-760-9502  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1091 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-633-768A-2  
Query Match 7.7% Score 356; DB 3; Length 1091;  
Best Local Similarity 23.8%; Pred. No. 1.4e-25;  
Matches 197; Conservative 117; Mismatches 308; Indels 204; Gaps 45;  
Qy 125 DLVLTTHASPRFTVSRRTSG-----DTLFDTAPGLVFRDKYLEVTSAL----- 168  
Db 193 DLNVIIYGN-FKTRVTRKRGKIMENNEVPAGSLGKCRGLMFVDRLYGTALASVNEY 251  
Qy 169 ---PAGRASLYGLG-----EHTKSSPRL-RHNDSTFLMNADIGASYDVNL---Y 211  
Db 252 RNDPDRKEGFGYGAENCFEWFDEQNRNKYILERTGIAMTNYD-NYNTNQSDLIAPGY 310  
Qy 212 GSHF-PY--MDVRAPGTAHGVLILLSSNGMDVLYGGSYV-----TYKVGCV----- 254  
Db 311 PSDNFVTPMYPAPWV---VVKCSGNSDEQYSGWFMNDVVSQTYMNTGCTSMNCGEEN 367  
Qy 255 -----LDYFFAGPNPLAVVDQVT-----OLIARPAM-PYWSFGFHQC 292  
Db 368 LAYMGAQCGPFDQHFYVYGDGLEDVVQAFSLQKGEFENQVNLKRAVMPKPYVFGYFQG 427  
Qy 293 RYGYL-----NVSDLRVVAARYAKARIPLEVMVMTDIDYMDGPKDFDLDRVNET 340  
Db 428 VFIASLLRQRPEGGNIS-VGEIVEGYGNSNPFLGLEAVDVMQODLRVFTTKIEFTI 486  
Qy 341 AAELRPFVRLHRNAQKYLILDPG-----IRVDFIDATYGTFFVGMQODIFLKR 390  
Db 487 ANKVTGCD--SNNKSVFEWAHDKGLVCQNTVCTFLRNDNGGADYEVNQTLEKGLYTKN 544  
Qy 391 N---GTNP-VGNWPGDYV-----FDPMPHAAEFWAREIS-LFRRTIP 430  
Db 545 DSLTNTNFTGTTNDGFSDAYIGHLDYGGGNCDAFPDWPGEPPGVAEWMGDNYSKLEK--IG 602  
Qy 431 VDGLWIDMNEISNFYNPPEM-----NALDD-PPY---RINNDGTGRPINNKTVRPLAVHY 481  
Db 603 LDFVQDMTV-----PAMPHKVGDAVOTSPYGWENENDPSNGR-YNWKSYHPQVL-- 653  
Qy 482 GGVTE--YEEH-----NLFGLEA-RATGRGVLRDT---GRPPFVLSRSTFVGS 524  
Db 654 --VTDMRYENHGREPMFTQRNMHAYTLCESTRKEGIVANADTLTKFRSYYIISRGGYIGN 711  
Qy 525 GRYTAYMTGDNAATMGDLRYSLNTMLSPGLFGWPMIGADICGFGNTTEELCG-----RW 579  
Db 712 QHFGMWVGDNSSQRYLOMMIANIVNNMNSCLPLVGSIDGGFTSYDGRNVCFGDLWVRP 771  
Qy 580 IQLGAFYPPFSRDHSAIFT-----VRRRLYLPFSSVAASGRKALGLRYQLLPYFTILMYE 632  
Db 772 VOAGCLLPWFNHYGRVLEGEKQEGKYQELMYVYKDEMATLTKFIEFRYRQWELVLTAMYQ 831  
Qy 633 AHMTGAPIARELPFSYPHDVATYV-DROFLL-----GRGVLSVPVLEPGPTTVDAYFA- 686  
Db 832 NAAFGKPIIKAAAM-YNDRNRVGAQDDHFLGGHDGYRILCAPVWMENTTSRDLVPLV 890  
Qy 687 GRWYRL---YDYS-LAVATRTGKHVR-LPAPADTVNHLTGTTLPLOQSALTTSRART 741  
Db 891 TWTYKFGPDYTKRLDSALDGGQMKYNSVPSQSDSPFVREGALLPTRYTLDSGNSKMYT 950  
Qy 742 -----AFHLLVALAEDGTASGYLFLDDG-----DSPEYGRS 773  
Db 951 YTDKDPVFEVFLP-----GNNRADGCMGYLDDGGITTTDAEDHGKFS 991



## RESULT 10

```

US-08-633-768A-1
; Sequence 1, Application US/08633768A
; Patent No. 6013504
; GENERAL INFORMATION:
; APPLICANT: YU, SHUKUN
; APPLICANT: BOJSEN, KARSTEN
; APPLICANT: KRAGH, KARSTEN
; APPLICANT: BOJKO, MAJA
; APPLICANT: NIELSEN, JOHN
; APPLICANT: MARCUSSEN, JAN
; TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM
; TITLE OF INVENTION: A FUNGUS INFECTED ALGAE, ITS PURIFICATION
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/633,768A
; FILING DATE: 02-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 9321301.5
; FILING DATE: 15-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Altman, Daniel E
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: DY0U7.001APC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714-760-0404
; TELEFAX: 714-760-9502
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1088 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-633-768A-1

Query Match 7.5%; Score 349; DB 3; Length 1088;
Best Local Similarity 21.1%; Pred. No. 6.7e-25;
Matches 196; Conservative 139; Mismatches 308; Indels 284; Gaps 45;

QY 125 DLVLTVHASPRFTVSRSTGDTLFDT-----APGLVFRDK-YLEVTSA----- 168
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 192 DLSVLIYGN-FKTRVTRKSDGKVMENDEVTASSGNKRCGLMFVDRLYGLAIASVKNKF 250
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 169 ---PAGRASLYGLGHTKSSFLRHNSFTLWNADIGA---SYVDVN----- 209
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 251 RNDAVKQEGFYGAGE-----VNCKYQDTYILERTGIAMTNYNDLNYNQWDLRPPHHDG 305
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 210 -----LYGSHPFYMDVRAPGTAH-----GVLLSSNGMDVLYG 242
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 306 ALNPDIYTPMYAARPLVNGCAGTSEQSYGWPMDNVQSVMYMTGDTTWSNGEDLAYM 365
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 243 GSYTYTKVIGVLD--FYFFAGNPPLAVVDQYT-----QLIARPAPM-PYWSFGFH 290
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 366 GAQY-----GPFDOHFYVAGGMECVTAFSLLOGKEFENQVLNKSVMPPKYVGF 419
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 291 QCRVGYL-----NVSDLERVVARAKARIPLEVMTDIDYMDGFKOTLDRVN 338
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 420 QGVFGTSLLRAMPAGENNIS-VEEIVEGYQNNFPFEGLAVDVDMQDNLRVFTTKGEF 478
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY 339 FTAELRPVDELHRAOKYV-----LILDPGI-----RVDPIDATYGTFFVRGMQODIFLKR 390
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 479 WTANRVGTGGDPNNRSVFWEAHDKGLVCQTITCLFRDNBEGQDYEVNQTLERQLYTKN 538
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 391 N---GTNF-----VGNVWPG-----DVYFPDFWHPAFAAEFWAREI-SLFERRTP 430
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 539 DSLTGTGDMTDDGSPDAYIGHLDYGGGVECDALFPDNGRPDVAEMWGNKYKLF--SIG 596
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 431 VDGLWDM-----NEISNFYNPEP-----MNALDDPPYRINNDGTGRPINNKTVRPLAV 479
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 597 LDFVQDMTVPAMPHKIGDDINVKPDGNWPN-A-DP-----SNQYMKWTKYHPQVL 647
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 480 HYGGYTE--YEEH-----NLFGLEEARAGR-GVLRDTG-----RRPEVLSRSFTV 522
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 648 -----VDMRYENHGREPMVTQRNIHAYTLCTSRKEGIVENADTLTKFRSIIISRGYI 703
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 523 GSGRYTAYWTGDNAATWGDRLYSINTMLSFGLFGMPMIGADICGF-----NGN-----TTEE 574
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 704 GNOHFGMMVGDNSTTSNYIGMMIANNINMMNSCLPLVGSIDIGGFTSYDNENQRTPTG 763
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 575 LGRMIQLGAFYFP-----SRDHSALTFTVRELYLWPSVAASGRKALGLRYQLLP 624
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 764 LMVRYVQAGCLLPWFRNHYDRMIESKDHGKDY---QELYMYENEMDTLRKFVEFRYRGE 820
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 625 YFYLWYEAHMTGAPIARPLPFSYPHDVATYGVDRQFLL---GRGLVSPVLEPGPTTV 680
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 821 VLYTAMYQNAAFKGLIIKAAASMYNDSNVRRQRNDHFLGCHDGYRILCAPVWENSTER 880
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 681 DAYFPA-GRWY-----RLYDYSLAVATRTGKHVRLPAPADTWNVHL 720
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 881 ELYLPVLTQWYKFGPDFDTKPLEGAMNGGDIYNY-----PVPQSESPIFV 926
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 721 TCGTILPQOQSALTTSRART-----AFHLLVALAEDGTASGYLFLDDGDSPEYG 770
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 927 REGAILPTRYTLNGENKSLNTYTDPLVFEVFPPL-----GNRRADGMCYLDGCG----- 976
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 771 RRSKSMVRFNKYIPNNKGAIKVSEVHVNSVAQSRITLVISK-----VVLWGH 818
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 977 -----VTTNAEDNGKFSVVAAB-----QDGGTETITFTNDCEYVFGGFYVRVGA 1025
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 819 RSPAAPKKLTVHVNSAEVSAESSAGTR 845
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1026 QSPS-----NIHV-----SSGAGSQ 1040
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
US-09-275-608-4
; Sequence 4, Application US/09275608
; Patent No. 6541237
; GENERAL INFORMATION:
; APPLICANT: YU, Shukun
; APPLICANT: JSEN, Kirsten
; APPLICANT: MARCUSSEN, Jan
; TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE AND
; TITLE OF INVENTION: ITS USE IN THE PRODUCTION OF 1,5-ANHYDROFRUC
; TITLE OF INVENTION: TOSE
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/275,608
; FILING DATE:

```

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP95/02172  
FILING DATE: 06-JUN-1995  
APPLICATION NUMBER: 08/836,156  
FILING DATE: 15-APR-1997  
APPLICATION NUMBER: PCT/EP94/03397  
FILING DATE: 15-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Altman, Daniel E  
REGISTRATION NUMBER: 34,115  
REFERENCE/DOCKET NUMBER: DYOU9.001C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 714-760-0404  
TELEFAX: 714-760-9502  
TELEX:  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 570 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-275-608-4

Query Match 6.0%; Score 279; DB 4; Length 570;  
Best Local Similarity 21.4%; Pred. No. 1.8e-18;  
Matches 143; Conservative 87; Mismatches 226; Indels 212; Gaps 28;

QY 83 DHRPWEVDIIIPRPGDVLHDPASSAPLQGRVLSGASDLVLTVAHSPRFTVSR 142  
DB 7 DNLNYP-DLIP---PG---HSDPDYIIPM-----YFAAPWVIAHGYR 44  
QY 143 STGDTLFTDAPGLVFRDKYLEVTSALPAGRASLYGLGHTKSSPRLRHNDSTFLWNADIG 202  
DB 45 GTSD-----QISYGFNLONVSQSYNTGDDAW 71  
QY 203 ASYVDNLYGSHPFYMDVRAPGTAHGYLLSSGMDVLYGGSYVTVKYGVLDFYFAG 262  
DB 72 AGQKDLAYMGAQ-----CGFFDQHFVYEAAGDGLDVVTAFSYLOK----- 112  
QY 263 PNPLAVVDQYTO---LIARPAFMPYWSFGFHOCRYGL-----NVSDLERVYAR 308  
DB 113 -----EYENQNLNRSAMPKRYVFGFGVFGATSLRLDNLPAGENNVLS-LEEIVEG 163  
QY 309 YAKARIPILEVWMTDIDYMDGFKDFTLDRVNFATAELRPFVDRHLHNAQKYVLLDPGIRV 368  
DB 164 YQNQVFEGLAVDVDMQDLIRVFTTPAPFTANKVGEKGGDPNNKSVFEWAHDSGLVCQT 223  
QY 342 -----AELRPF-VDRHLHNAQKYVLLDPGIRVDPID-ATYGTFFVRGMOQDIFLKRNG 392  
DB 224 NVTCLKNKXNPYEVNQSLREKQLYT-----KSDSLDNI DFGTTPDGPSPD----- 268  
QY 393 TNFVGNWPG-----DYVFDPMHAPAAEFWAEI-SLPRFTIPVGLWIDMNEISNFYN 446  
DB 269 -AYIGHLDYGGVCECDALFPDWRGPDVAQWMDNYKKLF--SIGLDFVWQDMTV----- 319  
QY 447 PEPNALDDPPYRI-----NNDGTGRP-----INNKTVRPLAV-----HYG 482  
DB 320 PAMM-----PHRLCDPVGTNSSETAPGWNDKDPNSGRVNWKSYHPQVLVDMRVDDYG 373  
QY 483 GVTEYEHNLFGLEABATGR-GVLRDTG-----RRPFVLSRSTFVSGRYTAYWTGD-NAATWGDRLYS 545  
DB 374 RDPVITQNLHAYTLCBSTRREGIVGNADSLTKFRRSYIISRGYIGNHQHFQGGWMDGNS 433  
QY 537 ATWGDRLYSINTMLSFGLFPMGPMGADICGNGN-----TTEELCGRWIOLGAFYFSPRD 591  
DB 434 STEDYLAAMVINVINMMNSGVPLVDSIGDGGTTHDKENPCTPDLMRWFVQAGCLLFWFRN 493  
QY 592 HSAIFTVRR-----ELYLWPSVAASGRKALGLRYQLLPYFTLMVEAHMTGAPIARPL 644  
DB 494 HYDRWIESKKHGKQYQELMYTRDHLDAIRSFVELRYRWQEVLYTAMYQNALNGKPIIKTV 553

QY 645 FFSYPHDV 652  
DB 554 SM-YNNDM 560

RESULT 12  
US-09-634-238-331  
Sequence 331, Application US/09634238  
Patent No. 6544772  
GENERAL INFORMATION:  
APPLICANT: Glenn, Matthew  
APPLICANT: Havukkala, Ilkka J.  
APPLICANT: Bloksberg, Leonard, N.  
APPLICANT: Lubbers, Mark W.  
APPLICANT: Dekker, James  
APPLICANT: Christensson, Anna C.  
APPLICANT: Holland, Ross  
APPLICANT: O'Toole, Paul W.  
APPLICANT: Reid, Julian R.  
APPLICANT: Coolbear, Timothy  
TITLE OF INVENTION: Polynucleotides, materials incorporating  
TITLE OF INVENTION: them and methods for using them.  
FILE REFERENCE: 11000.1043UI  
CURRENT APPLICATION NUMBER: US/09/634,238  
CURRENT FILING DATE: 2000-08-08  
NUMBER OF SEQ ID NOS: 422  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 331  
LENGTH: 390  
TYPE: PRT  
ORGANISM: Lactobacillus rhamnosus  
US-09-634-238-331

Query Match 4.3%; Score 197; DB 4; Length 390;  
Best Local Similarity 24.6%; Pred. No. 1.2e-10;  
Matches 95; Conservative 35; Mismatches 166; Indels 90; Gaps 16;

QY 309 YAKARIPILEVWMTDIDYMDGFKDFTLDRVNFATAELRPFVDRHLHNAQKYVLLDPGIRV 368  
DB 42 YQPGKIPAG-----QSGIKESLNGELNNYQFSARAVIDRYQKNDMPGLGWL----- 87  
QY 369 DPIDATYGTFFVRGMOQDIFLKRNGTNFVGNWPGDVYYPDF--MHPAAAEFWARISLFR 426  
DB 88 -PNDCG-YGA---GYQTDITLGNLQNLKS-----PADYADQHGCVATGLWTOO----- 129  
QY 427 RTPVDGLWIDMNEISNFYNPEPNAALDDPPYRINNDDGTGRPINNKTVRPLAVHYGGYTE 486  
DB 130 NLSPPDPA-----NPKP-----DORDF-AKEVAIG-----VYALKTDVAVWGS 166  
QY 487 YEEHNLFGLEABATGRGLRDTGRRPFVLSRSTFVSGRYTAYWTGD-NAATWGDRLYS 545  
DB 167 GYSFGLDGLAKADAMMTQVKGDS-LRPFPAITLDGWAGTQRYAGVWGTGDTGQGWEIFRHF 225  
QY 546 INTMLSFGLFPMGPMGADICGNGNNTTEELCGRWIOLGAF-----YFSPR 590  
DB 226 IPTYGTGLSGQYVGSMDGIFGGGNFVNTDRFQWKAFTPQLNMDGNGANPKTPESF 285  
QY 591 HSAIFTVRRRELYWPSVAASGRKALGLRYQLLPYFTLMVEAHMTGAPIARPLFFSYPH 650  
DB 286 DQ-----QTTAINRANKOKTMLMPPNYTASQSVDPDGKPMWRGLFLDIPN 331  
QY 651 DVATYTG--VDRQFLGLGRGLVSPVLE 674  
DB 332 IPEAVTDLVKVEYLWGNDFLVAPIYQ 357

RESULT 13  
US-09-437-054A-10  
Sequence 10, Application US/09437054A  
Patent No. 6316698  
GENERAL INFORMATION:  
APPLICANT: Allen, Stephen M.  
APPLICANT: Kinney, Anthony J.

;; TITLE OF INVENTION: Plant Alpha-Glucosidase II Homologs

;; FILE REFERENCE: BBI273 US NA

;; CURRENT APPLICATION NUMBER: US/09/437.054A

;; CURRENT FILING DATE: 2001-05-14

;; PRIOR APPLICATION NUMBER: 60/107,909

;; PRIOR FILING DATE: 1998-No. 6316698ember-10

;; NUMBER OF SEQ ID NOS: 19

;; SOFTWARE: Microsoft Office 97

;; SEQ ID NO 10

;; LENGTH: 150

;; TYPE: PRT

;; ORGANISM: Triticum aestivum

;; FEATURE:

;; NAME/KEY: UNSURE

;; LOCATION: (73)..(74)

;; NAME/KEY: UNSURE

;; LOCATION: (102)

;; NAME/KEY: UNSURE

;; LOCATION: (123)

;; NAME/KEY: UNSURE

;; LOCATION: (141)

;; NAME/KEY: UNSURE

;; LOCATION: (143)

US-09-437-054A-10

Query Match 4.2%; Score 195.5; DB 4; Length 150;

Best Local Similarity 30.4%; Pred. No. 3.2e-11;

Matches 49; Conservative 27; Mismatches 62; Indels 23; Gaps 6;

QY 615 ALGRLYQLLYFYILMYEAMHTGAPIAPLPFSYPHVDVATYGVDRQFLLRGVVSPVLE 674

DB 1 AIHMRYSLPYYSLSFQESATYGVVMPLEPLWLEFPDDKETYNNGEAPMVGPSILAQGIYE 60

QY 675 PGPTTYDAYFAGR-WRLYD---YSLAVATRTGKVRLEPAPATVNVHLTGGTILPQQ 730

DB 61 EGQKSVSVLPQXLMYOLRNGSPYKGSV-----H-KLQVSEDSIPXFORSGTIVPRK- 113

QY 731 SALTTSRARR-----TAFHLLVALABDGTASGVLFLLDDG 764

DB 114 -----DRFRILTOXVNDSTYTLVIGLNNNSWLVXXXTYVDDG 149

RESULT 14

US-09-410-551B-72

;; Sequence 72, Application US/09410551B

;; Patent No. 6503737

;; GENERAL INFORMATION:

;; APPLICANT: KOSAN BIOSCIENCES, Inc.

;; APPLICANT: REEVES, CHRISTOPHER

;; APPLICANT: CHU, DANIEL

;; APPLICANT: KHOSLA, CHAITAN

;; APPLICANT: SANTI, DANIEL

;; APPLICANT: WU, KAI

;; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA

;; FILE REFERENCE: 30062-20026.00

;; CURRENT APPLICATION NUMBER: US/09/410.551B

;; CURRENT FILING DATE: 1999-10-01

;; PRIOR APPLICATION NUMBER: US 60/139,650

;; PRIOR FILING DATE: 1999-06-17

;; PRIOR APPLICATION NUMBER: US 60/123,810

;; PRIOR FILING DATE: 1999-03-11

;; PRIOR APPLICATION NUMBER: US 60/102,748

;; PRIOR FILING DATE: 1998-10-02

;; NUMBER OF SEQ ID NOS: 72

;; SOFTWARE: FastSeq for Windows Version 4.0

;; SEQ ID NO 72

;; LENGTH: 6396

;; TYPE: PRT

;; ORGANISM: Streptomyces hygrosopicus

US-09-410-551B-72

Query Match 2.6%; Score 122.5; DB 4; Length 6396;

Best Local Similarity 20.4%; Pred. No. 0.4;

Matches 186; Conservative 73; Mismatches 294; Indels 357; Gaps 45;

QY 16 FAPRLCSSKEGFLAARTVLAIVTMEGALRAEATGRRSSTGVDYORLAVYASLETDS-- 73

DB 3336 FAQRLAELPDADRGAALTTLVSDAT-----AAVLGHADASEIAPTPTTKDKLGIDSLT 3387

QY 74 --RLRVITDA-----DHP-----RWEVQDIIIPRA-PGDVLHDAPPA- 109

DB 3388 AIELRNRLAEATGLRLSATLVFEDHPTPRVLAAGKLDLFGTAIVTPPARTARTHHDEPLAI 3447

QY 110 --SSAPLOQGRVLSF-----AGSDLVLTVHASPRFTVRRSSTGDTLFD---TAPGLV 156

DB 3448 VGMACELPGVASPEDLWQLVASGTDAT-----EFTDRGWDIDRLFDPPDPAPGKT 3500

QY 157 F--RDKYLEVTSALPAG-RASLYGLGEHTKSFRLRHNDSP-TLWNADIGASYVDNLYG 212

DB 3501 YVRHGGFL-----AEAAGFDAAFFGISPREARAMDQQRVILLETSEAFENAGIVPDTLRG 3556

QY 213 SHP-FYMDVRAEGTAHGVLLSSNGMDVLYGGSYTYKYVIGGVLDYFFAAGNPLAVVD- 270

DB 3557 SDTGVMGAFSHGYGAGVDL---GG---FGATATQNSVLGRLS-YFFGMEGPAVTVD 3608

QY 271 -----QYTQLI-----ARPAPMYMSFGFHQCRYGL----- 297

DB 3609 ACSSSILVALHQAALRTGECSLALAGGVYVMTPLGYVEF---CRQGLAPDGRQCAF 3664

QY 298 -----NVSDLERVVARYAKARIPLEVWMTDIDYMDGKDFTLDRVNTAAEL 344

DB 3665 AEGADGTSFEGAGVILVERLSDAERNHGTVLAVRSSAVNQDGAS-----NGISAPN 3717

QY 345 REPVDRLHNAQKYVILIDPGIRV-----DPID-----ATYG----- 376

DB 3718 GPSQQRVIRQALDKAGLAPADVVDVVEHGTGTPGDIQAQAIATYQODRTPYLGSV 3777

QY 377 -----TFVRGMOQDIFLRKNGTFNFGVNWPGDYY----- 405

DB 3778 KSNIGHTQTAGVAGVIXVMAMRHG-----IAPKTLHVDEPSSHVDWTEGAVELLTEA 3831

QY 406 --FPDFMHFAAEFWAREIS-----LFRRTIP-----VDGLMIDMNEISFNYPNPEM 450

DB 3832 RPWPDAGRPRRAGVSLGISGTNAHVILEGVPGRSVPESVDGL-----VPLEV 3880

QY 451 NALDDPPYRINDGTGRPINNKTVRPLAVHYGVTEYBEHNLFGLEABATGRGVLRDTG 510

DB 3881 SARSEASLR-----GOVERLEGYLRGSVDVAVAQGLVRE-- 3915

QY 511 RRPFFVLSRSTFVSGRYTAYWTGDNAATWGLDLR-YSINTMLSFGLFMPMPMGADICGFNG 569

DB 3916 -----RAVF-----GHRVLLGDARVMGVAVDQPRTVFVFPQCGA----- 3950

QY 570 NNTBELCGRMIQLGAFYFESRDSALFTVR-----RELYLWPSVA----- 609

DB 3951 -----QWVGWGVEL---MDRSVFAARMECARALLPHTGWDVREMLARPDAERVE 3999

QY 610 ----ASGRKALGLRYQLLPYFTVLMYEAH-----MTGAPLARPLFFSYPHD 651

DB 4000 VVQPASWAVAVSL-----AALMQAGVVDVAVIGHSQGEIAAACVAGALS---ED 4047

QY 652 VATYGVDRQ-----FLLRGVLVSPVLEPGTPTVDAYFPAGRWYRLYDYSLAVATRTGKH 706

DB 4048 AARVVALRSQVIAARLAGRGAMASVALPAG---EVGLVEGVW-----IAARNG--- 4092

QY 707 VRLPAPADTV 716

DB 4093 -----PASTV 4097

RESULT 15

US-09-198-452A-966

;; Sequence 966, Application US/09198452A

;; Patent No. 6559294

;; GENERAL INFORMATION:

```

; APPLICANT: Griffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 966
; LENGTH: 692
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-966

```

```

Query Match      2.6%  Score 121.5;  DB 4;  Length 692;
Best Local Similarity 18.9%;  Pred. No. 0.0099;
Matches 140;  Conservative 85;  Mismatches 209;  Indels 305;  Gaps 39;

Qy 168 LPAG-----RASLYGLGEHTKSSFLRHNSFTLW-NAD-----IGASYVDVNLG----- 212
Db 54 LPLGCAELAAVLG-----YVLRQNPDPHWRNDRFVLSAGHSALLYSCLHLAGF 105

Qy 213 -----SHIPFMDV-----RAPGTAHGVLILSSN----- 235
Db 106 DVSLEDLQEFRLQHSRTPGHPEYGETVGEVATTGPGQLGNAGVGMALSMKMLESRFNR 165

Qy 236 GMDVLYGGSY-----VTYKVIQGVLDFFVFAGP---NPLAVWDQYTOTLIARPAP 281
Db 166 GHEIFNGKIYCLAGDGCFCMEGVSHVCS-----FAGSLNLANLVVIYDYNVVDL--- 215

Qy 282 MPYMSFGHCRCRYGLNVSDLERVARYAKARIPLEVMWTDIDYMGFKDFTLDRVNFYA 341
Db 216 -----GYLNEISVEDTKRF-----EAYGWYVEIDGY-DFTIHETFS 254

Qy 342 AELRPFPVDRLHRNAQKVL-----ILDPIRVDPIDATYGTFFVRGMOODIFLKRNGTNFV 396
Db 255 -----IKRQERPLVIAHTIIGHGSPKGTNKAHGSPL-GVE-----GTHET 296

Qy 397 GNWFGDVYFPD---FMHPAAAEFWAREISLFRRTIPVDGLWIDMNEISNFNPE----- 448
Db 297 KQFW---HLPEEKFFVPPAVKQFFAHKIQEDRK---AQEQWLDEVRVWSKQFPFLHEEF 349

Qy 449 -----PMN-----ALDDPPYRINNDGTGRPINNKTVRPLAVH-----YGVTEYEEH 490
Db 350 VALTSHKLPKNLSLVQSVEMP-----DSIAGRAASNKLIQVLVQHIFPYLIGGSADLSS- 403

Qy 491 NLFGLLEBARATGRGLVDRDTCRRPFLSRSTFVSGSR---YTAYWTGONAATWGLDRYSINT 548
Db 404 -----SDGTWIANEKVIHTYDFSGRN-----IKYGVRE 431

Qy 549 MLSFGLFGMPMIGADICGFNGNTTEELCGRWIQLGAFYFESRDHSAIFTVRRELYLWPSV 608
Db 432 -----FGMATI-----MNGLAYSQV-----PRPF-----GGTFLV----- 456

Qy 609 AASGRKALGRYQLLPYFTLMYEAHTGAPIARPLFFSYPHDVATYVDVDRQELLGRGVL 668
Db 457 -----FSDYMRNAIRLAALSCLPVIYQFTHDSIFVG----- 487

Qy 669 VSPVLEPGPTTVDAYFPAGRWYRLDYSLAVATRTGKHVRLPAPADTV-----NVHLTG 722
Db 488 -----EDGPT-----HQPVEQLMSLRALPGLYVIRPADANEVRGANIAGLKHTG 531

Qy 723 GTILPQOQSALTTSRARTAFHLLVALAEDGTASG-YLFL-DDGDSPEYGRSDWSMYRF 780
Db 532 PTVIVLSRQALPTLPAHRPF-----KDGVGGRGAYIVLKESEKPDYTLFATGSEVSL 584

Qy 781 NYKIPNNKGAIVKSEVVH 799
Db 585 -----ALSVAKELEH 594

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 : Search time 47.3131 Seconds  
(without alignments)  
2942.169 Million cell updates/sec

Title: US-10-043-418-1

Perfect score: 4630

Sequence: 1 MATVGVLLCLCLFAPRL.....IGGLSLVGEPELKVAMSY 877

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_19Jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	4630	100.0	877	19	AAW59040 Barley alpha-glucosidase
2	2324.5	50.2	901	23	AAU97731 Common buckwheat alpha-glucosidase
3	2324.5	50.2	901	23	ABR09151 Common buckwheat alpha-glucosidase
4	2297.5	49.6	901	23	AAU97732 Common buckwheat alpha-glucosidase
5	2297.5	49.6	901	23	ABR09152 Common buckwheat alpha-glucosidase
6	1993.5	43.1	682	21	AAV51670 Potato alpha-glucosidase
7	1410.5	30.5	985	14	AAK42995 Glycosyltransferase
8	1410.5	30.5	985	14	AAK42214 Aspergillus niger
9	1407.5	30.4	922	10	AAW88044 An enzyme with sug

10	1403	30.3	953	23	ABBS7174	Mouse ischaemic co
11	1348.5	29.1	985	18	AAW15191	Aspergillus oryzae
12	1329.5	28.7	1827	20	AAW74090	Human hsi protein
13	1329.5	28.7	1827	22	AAU09028	Human sucrase isom
14	1329.5	28.7	1829	22	ABG14994	Novel human diago
15	1328	28.7	963	20	AAV49895	Endomyces fibulige
16	1279.5	27.6	958	9	AAAP81181	Sequence of glucoc
17	1279.5	27.6	958	11	AAAR07575	Glucosylase encod
18	1141	24.6	1070	13	AAAR21521	Alpha galactosidas
19	963	20.8	914	23	ABP52437	Human carbohydrate
20	956	20.6	912	23	ABG69613	Human NOV11d prote
21	953.5	20.6	914	23	ABG69611	Human NOV11b prote
22	952.5	20.6	914	23	ABG69612	Human NOV11c prote
23	939	20.3	912	24	AAAG79779	Carbohydrate-assoc
24	913.5	19.7	941	22	AAAM80202	Human protein SEQ
25	913.5	19.7	944	22	AAAT79218	Human protein SEQ
26	900	19.4	967	23	ABG69610	Human NOV11a prote
27	883	19.1	914	23	AAAG66094	Soybean alpha-gluc
28	826.5	17.9	919	18	AAAM18580	Potato alpha-gluc
29	826.5	17.9	919	18	AAAG66099	S. tuberosum alpha
30	818	17.7	565	21	AAAS88449	Breast and ovarian
31	809.5	17.5	924	22	ABBS7421	Drosophila melanog
32	809.5	17.5	924	22	ABBS7429	Drosophila melanog
33	801	17.3	763	23	ABBS01119	Listeria monocytog
34	724	15.6	444	21	AAAB42826	Human ORFX ORF2590
35	679.5	14.7	693	24	ABP96604	S. solfataricus ma
36	679.5	14.7	712	24	ABP96616	Plasmid pNOV4839 m
37	679.5	14.7	718	24	ABP96615	Plasmid pNOV4831 m
38	679.5	14.7	718	24	ABP96625	MalA fusion protei
39	625	13.5	834	11	AAAR04869	Alpha-1-6-glucanas
40	595.5	12.9	390	23	ABP51380	Human WDDT SEQ ID
41	564	12.2	331	22	AAU23490	Novel human enzyme
42	553	11.9	359	22	AAU19420	Human diagnostic a
43	541.5	11.7	235	22	AAAM00087	Alpha-glucosidase
44	466	10.1	1066	16	AAAT70638	M. costata alpha-1,
45	466	10.1	1066	16	AAAT72711	Alpha-1,4-glucan 1

#### ALIGNMENTS

RESULT 1  
AAW59040  
ID AAW59040 standard; Protein; 877 AA.  
XX  
AC AAW59040;  
XX  
DT 07-AUG-1998 (first entry)  
XX  
DE Barley alpha-glucosidase protein.  
XX  
KW Alpha-glucosidase; barley; starch grain hydrolysis; alpha-amylase;  
KW glucoamylase; industry; germplasm; hydrolytic enzyme.  
XX  
OS Hordeum vulgare.  
XX  
PN US5763252-A.  
XX  
PD 09-JUN-1998.  
XX  
PF 28-APR-1995; 95US-0430925.  
XX  
PR 28-APR-1995; 95US-0430925.  
XX  
PA (WISC ) WISCONSIN ALUMNI RES FOUND.  
XX  
PI Skadsen RW, Tibbot BK;  
XX  
DR WPI; 1998-347329/30.  
XX  
N-PSDB; AAV11736.  
XX  
PT DNA encoding barley alpha-glucosidase protein - useful for producing recombinant protein to increase rate of starch grain hydrolysis when

used with alpha amylase

Claim 1; Col 21-26; 19pp; English.

This sequence represents a novel barley alpha-glucosidase protein. Recombinant alpha-glucosidase can be used to increase the rate of starch grain hydrolysis when used together with alpha-amylase or can supplement glucosylase in industrial starch hydrolysis systems. Useful DNA sequence characteristics from this enzyme can be identified which can be used as hybridisation probes for identifying germplasm with high levels of efficient hydrolytic enzymes.

Sequence 877 AA;

Query Match 100.0%; Score 4630; DB 19; Length 877;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 877; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MATVGVLCLCLCLFAPRLCSKEEGPLAARTVLAVAVTMEGALRAEAATGGRSSTGDV 60  
 1 MATVGVLCLCLCLFAPRLCSKEEGPLAARTVLAVAVTMEGALRAEAATGGRSSTGDV 60

61 QRLAVASLETDSRLVRITDADHPREVPODIIIPRAPGDLVHDAPPASSAPLQGRVLS 120  
 61 QRLAVASLETDSRLVRITDADHPREVPODIIIPRAPGDLVHDAPPASSAPLQGRVLS 120

121 PAGSDLVLTVHASPRFTVSRRSTGDTLFDTPAGLVFRDKYLEVTSALPAGRASLYLGE 180  
 121 PAGSDLVLTVHASPRFTVSRRSTGDTLFDTPAGLVFRDKYLEVTSALPAGRASLYLGE 180

181 HTKSSFLRHNDSETLMNADIGASYVDVNLGSHPFYMDVRAPGTAHGVLLLSNGMDVL 240  
 181 HTKSSFLRHNDSETLMNADIGASYVDVNLGSHPFYMDVRAPGTAHGVLLLSNGMDVL 240

241 YGGSVTVTKVGGVLDFFVFAGPPLAVVDQYTLIARPAAMPYWSFGHCRCRYGLNVS 300  
 241 YGGSVTVTKVGGVLDFFVFAGPPLAVVDQYTLIARPAAMPYWSFGHCRCRYGLNVS 300

301 DLERVARYAKARIPLEVMTDIDYMDGFKDFTLDRVNFATAELRPVDRLHRNAQKYL 360  
 301 DLERVARYAKARIPLEVMTDIDYMDGFKDFTLDRVNFATAELRPVDRLHRNAQKYL 360

361 ILDFGIRVDPIDATYGTFRVGMQDIFLKRNGTNFVGNVWPGDYFFDFMHPAAAEFAR 420  
 361 ILDFGIRVDPIDATYGTFRVGMQDIFLKRNGTNFVGNVWPGDYFFDFMHPAAAEFAR 420

421 EISLFRRTIPVDGLWIDWNEISNFYNPEPMALDDPPYRINNDGTGRPINNKTVRPLAVH 480  
 421 EISLFRRTIPVDGLWIDWNEISNFYNPEPMALDDPPYRINNDGTGRPINNKTVRPLAVH 480

481 YGGVTEYEHNLFGLLEARATGRVLRDTRGRPFVLSRSTFVSGRYTAYWTGDNAAATWG 540  
 481 YGGVTEYEHNLFGLLEARATGRVLRDTRGRPFVLSRSTFVSGRYTAYWTGDNAAATWG 540

541 DLRSYINTLSFGFLGMPMIGADICGFNGNTTEELCGRWIQLGAPYFSDHSAIFTVR 600  
 541 DLRSYINTLSFGFLGMPMIGADICGFNGNTTEELCGRWIQLGAPYFSDHSAIFTVR 600

601 ELYLWPSVAASGRKALGRYOLLPYFYTLMEYAHMTGAPIAPLFFSPYHDVATYGDQR 660  
 601 ELYLWPSVAASGRKALGRYOLLPYFYTLMEYAHMTGAPIAPLFFSPYHDVATYGDQR 660

661 FLLEGVLVSPVLEPGPTTVDAYFAGRWRYLYDYSLAVATRTGKHLRPLAPATVNVHL 720  
 661 FLLEGVLVSPVLEPGPTTVDAYFAGRWRYLYDYSLAVATRTGKHLRPLAPATVNVHL 720

721 TGGTILPQQSALTTSRARRTAFHLLVALAEDGTASGLFLDDGDSPEYGRSDWSMVR 780  
 721 TGGTILPQQSALTTSRARRTAFHLLVALAEDGTASGLFLDDGDSPEYGRSDWSMVR 780

781 NYKIPNNKGAIKVSEVHNYSQAQRTLVISKVLMGHRSPAAPKCLTVHVNSEAVEASS 840  
 781 NYKIPNNKGAIKVSEVHNYSQAQRTLVISKVLMGHRSPAAPKCLTVHVNSEAVEASS 840

QY 841 SAGTRYQVAGGLGGVAHIGGLSLVVGSEFELKVAMSY 877  
 DB 841 SAGTRYQVAGGLGGVAHIGGLSLVVGSEFELKVAMSY 877

RESULT 2

AAU97731

ID AAU97731 standard; Protein; 901 AA.

XX AAU97731;

XX 23-AUG-2002 (first entry)

XX Common buckwheat alpha-glucosidase polypeptide #1.

XX Common buckwheat; alpha-glucosidase; enzyme;

KW Fagopyrum esculentum Moeench.

XX Fagopyrum esculentum.

PN JP2002065273-A.

XX 05-MAR-2002.

XX 31-AUG-2000; 2000JP-0262104.

XX 31-AUG-2000; 2000JP-0262104.

XX (NISO ) NIPPON SHOKUHN KAKO KK.

XX WPI: 2002-474198/51.

XX N-PSDB; ABK86269.

PT A recombinant vector containing alpha-glucosidase gene, and a

transformant useful for preparation of alpha-glucosidase -

XX Claim 2; Page 7-8; 18pp; Japanese.

XX The invention relates to a recombinant vector containing

alpha-glucosidase gene derived from buckwheat. A transformant containing

the recombinant vector can be used for preparing alpha-glucosidase by

culturing the transformant and isolating alpha-glucosidase from the

culture. This sequence represents a buckwheat alpha-glucosidase protein.

XX Sequence 901 AA;

Query Match 50.2%; Score 2324.5; DB 23; Length 901;

Best Local Similarity 53.2%; Pred. No. 4.8e-192;

Matches 478; Conservative 121; Mismatches 259; Indels 41; Gaps 18;

QY 7 LLLCLCLCLFAPRLCS---SKEEGPLAA---FTVLAVAVTMEGALRAEAATGGRST--G 58

DB 13 LLLLAATLUF----CSLFVVSSEGVGYGVYVRAKVDSSNTLTAFKLINASSLYGQ 68

QY 59 DVORLAVASLETDSRLVRITDADHPREVPODIIIPRAPGDLVHDAPPASSAPL--Q 115

DB 69 DIPNLFTTATFEKDYELRIRITDAEKPRWEIPEVLRDGS--HGHPDLSRPTTPS 125

QY 116 GRVLSPAGSDLVLTWH-ASPFRFTVSRRTGDTLFDTPA-----GLVFRDKYLEVTS 166

DB 126 AAVLTHPNSDLIFRLHDTNPFQFSVTRRTNDVLFDRSADPETDPVGLVFKDQYIQLSS 185

QY 167 ALPAGRASLYLGEHTKSSFLRHNDSETLMNADIGASYVDVNLGSHPFYMDVRAP-CT 225

DB 186 SUPGRRALHYGHEHTKPTFLRAHNQTLTMNADIASYNVDLNLGSHPFYLDVRAPLGT 245

QY 226 AHGVLLLSNGMDVLYGGSVTVTKVIGVLDFFVFAGPPLAVVDQYTLIARPAAMPY 285

DB 246 SNGVLLLSNGMDVEYTGDRITTKVIGGIIDLYVFAGPTPDEVVQYTELIGRPAMPY 305

QY 286 SFGFHQCRVYGLNVSLELVRVARYAKARIPLEVMTDIDYMDGFKDFTLDRVNFATAELR 345



Db 842 SRGQVKGVTM-LSIGRWTTTSVKGARKSRGTGKEDWVEIPNLSLLVGRNFKLDIQIT 899

RESULT 4  
AAU97732 ID AAU97732 standard; Protein; 901 AA.

XX AAU97732;  
XX 23-AUG-2002 (first entry)  
XX Common buckwheat alpha-glucosidase polypeptide #2.  
XX Common buckwheat; alpha-glucosidase; enzyme;  
KW Fagopyrum esculentum Moeench.  
XX Fagopyrum esculentum.  
OS JP2002065273-A.  
PN 05-MAR-2002.  
XX 31-AUG-2000; 2000JP-0262104.  
PF 31-AUG-2000; 2000JP-0262104.  
XX (NISO ) NIPPON SHOKUHN KAKO KK.  
PA WPI; 2002-474198/51.  
DR N-PSDB; ABK86270.  
XX A recombinant vector containing alpha-glucosidase gene, and a  
PT transformant useful for preparation of alpha-glucosidase -  
PS Claim 2; Page 8-9; 18pp; Japanese.  
XX The invention relates to a recombinant vector containing  
CC alpha-glucosidase gene derived from buckwheat. A transformant containing  
CC the recombinant vector can be used for preparing alpha-glucosidase by  
CC culturing the transformant and isolating alpha-glucosidase from the  
CC culture. This sequence represents a buckwheat alpha-glucosidase protein.  
XX

Query Match 49.6%; Score 2297.5; DB 23; Length 901;  
Best Local Similarity 54.2%; Pred. No. 1.1e-189;  
Matches 455; Conservative 124; Mismatches 231; Indels 29; Gaps 14;

Qy 59 DVQRLAVVASLETDSRLRVIRITDADHPWEVPQDIIPRAPCDVLHDAPPASSAPL---Q 115  
Db 69 DIPNLTFATFEKDYRLIRITDAEKPRWEIPNEVLHRDGS---HGHQPLDSRPTTPS 125  
Qy 116 GRVLSAGSDLVLTWH-ASPRFTVSRKSTGDTLFDTPAP-----GLVFRDKYLEVTS 166  
Db 126 AAVLTHPNSDLIFRLHDTNPFGFSVTRRSRTNDVLFDTRSADPETDPVGLVFKDQYIQLSS 185  
Qy 167 ALPAGRASLYGLGHTKSSFLRHNDSTLWADIGASYVDNLYGSHPPFMDVRAP-GT 225  
Db 186 SLPADRSNLYGIGETKFTFLARNQTLTNWADIASYNVDNLYGSHPPFMDVRAPLGT 245  
Qy 226 AHGVLILLSNGMDVLYGGSYYTYKVGVLDFYFFAGFNPLAVVDQYTLQIARPAPMPY 285  
Db 246 SNGVLLNSNGMDVEYTKDKITYKVGIVDLYVFEGETPDEVVQYTELIGRAPMPY 305  
Qy 286 SFGFHQCRYGYLNSDLERVVARYANARIPLEVMMTDIDYMDGKDFTLDRVNTAAELR 345  
Db 306 SFGFHQCRYEYRNISSWENVVKAYSTMRIPLEAMWTDIDYMEANKDFTVDPVNFPLDKMQ 365  
Qy 346 PFVDFLHRNAOKYVILDPGRVDPIDATYCTYFVRGQODIFLXKNCNTFNFGVNWPGD 405  
Db 366 RFVNKLHKNQKQYVAILDPGININ--TTTYGTFORAMKADIFIKRQSEPYQGEVPG 423  
Qy 406 PFDGHPAAAEFWAREISLFRRTIPVDGLWIDMNEISNFYN--PEPMNALDDPPYRINND 463

Db 424 PFDLNPKTTFWISIEIQTFNALPVDGLWIDMNVSNFISFPDIPDPLNPPVNN 483  
Qy 464 GTGRPINNKTVRPLAVHYGGVTEYEENHLPGLLEAPATGRGLRDTGRPPFVLSRSTFVG 523  
Db 484 GGRRPINEKTIPIVSSVHYGNVSDYNNHNLGYLEATATNVALKKVTKQRPFFVLSRSTFIG 543  
Qy 524 SGRTYATWTGDNAATWGLDLRYISINTMLSFGLFGMPMIGADICGFNGNTTEELCGRWIQLG 583  
Db 544 SGKYTAHWTGDNAATWTNMAASIPILDFGLFGIPMIGADICGFAYTTEELCREWIQLG 603  
Qy 584 AFYFSDRHSALPTVRRRELYLWPSVAASGRKALGRYOLLPYFTLMTYEAHMTGAPIARP 643  
Db 604 AFYFSDRHSNDLSEPOELTQWGSVESARKVLGRYLLPYFTLMTYEAHMTGAPIARP 663  
Qy 644 LFFSYPHDVATYGVDRQFLGRVLVSPVLEPGPTTVDAYFPAGRWYLYDYSLAVAT-R 702  
Db 664 LFFSPNDPNTYIGIDAQFLIGKGVMSVPLTQCATSVTAYFPBGNWFLNFKYTKTVSSPA 723  
Qy 703 TGHVRLPAPADTVNVHLTGGTILPLQOSALTTSRARETAHLLVALAEDGTAS--GYLP 760  
Db 724 NGSFVTLDAPLKEINVHVRREGSILALQGEAMTTRARNTPELVVVVISDSNGSGSIGSVF 783  
Qy 761 LDDGDSPEYGRSD--WSMVRFNK-KIPNNKGAIKVSEVHNSYQAQSRTLVISKVVLGMH 818  
Db 784 LNVGVDIENGDDGRWSLVFTSAGLVGNK--VTITSSVNGRFPALSCQWKISKVTLGL 841  
Qy 819 RSPAAPKLTTHVNSAEVEASSAGTRYQNAQ-GLUGVAHIGGLSLVVGSEFELKVAMS 876  
Db 842 SRGQVKGVTM-LSIGRWTTTSKAGARKCKGTGKGVVEIPNLSLLVGRNFKLDIQIT 899

RESULT 5  
ABB09152 ID ABB09152 standard; Protein; 901 AA.  
XX ABB09152;  
XX 01-JUL-2002 (first entry)  
DE Buckwheat alpha-glucosidase protein SEQ ID NO:2.  
XX Buckwheat; alpha-glucosidase; enzyme; Fagopyrum esculentum Moeench;  
KW plant.  
XX Fagopyrum esculentum.  
FH Key Location/Qualifiers  
FT Peptide 1..31 /label= signal  
FT Protein 32..901 /label= alpha\_glucosidase  
XX JP2002065272-A.  
PD 05-MAR-2002.  
XX 31-AUG-2000; 2000JP-0262102.  
XX 31-AUG-2000; 2000JP-0262102.  
PR (NISO ) NIPPON SHOKUHN KAKO KK.  
XX WPI; 2002-346763/38.  
DR N-PSDB; ABU51447.  
PT A buckwheat-derived alpha-glucosidase gene -  
XX Claim 1; Page 9-10; 21pp; Japanese.  
XX The present sequence represents an alpha-glucosidase isolated from  
CC buckwheat (Fagopyrum esculentum Moeench). The alpha-glucosidase gene  
CC can be used for the preparation of alpha-glucosidase derived from





QY	522	VGSGRYTAYWTGDNAATGDLRYRSINTMLSFGLFGPMIGADICGPNNGNTTBELCGRWIQ	581
DB	331	LGSGRYTSHWTGDNAATWNDLAYSPTILSFGLFGIPMWGADICGFSNTTBELCRRMIQ	390
QY	582	LCAGFPFSDSHSAITVTRELYLWPSVAASGRKALGLRYQLLPYFYTLMYEAMTGPATIA	641
DB	391	LCAGFPFARDHSAKOTTPQELYSMDVSAAAAKVGLRLYQLLPYFYTLMYEAMTGPATIA	450
QY	642	RLPFTSYPHDVATYGVDRQFLLGRGVLSPVLEPGPTTVDAYFPAGRWRLYDYS LAVAT	701
DB	451	RELFTSFPODAKTFDISTQFLLGKGVMIISPILKOGATISVDAYFPAGWNFDLFNYSRVSU	510
QY	702	RTGKHVRLPAPADTVNVHLTGTITLPLQOSALUTTSRARTAFHLLVALAEDGTASGYLE	761
DB	511	NGQTYMTLDAAPPDHINHVHREGNILVMQGEAMTTQAAQRTAFKLLVVLSSKNSTGELFV	570
QY	762	DGDSPEYGRSD--WSMTRFNYKIPNNKGAIKYKSEVWHSYAGSRTLVLISKVLMGHS	820
DB	571	DDDEYQMGREGGRWTLTKFNSIIGNK--IVVKSEVWNGRYALDQGLVLEKVTLLGFEN	628
QY	821	PAAPKKLTVHNSAEVEASSAGTRYQ-----NAGGLGCGVAHIGGLSLWGWGEPELKV	873
DB	629	VRGLK-----SYELVGHQOQNTTMKESLKQSQGFVTWIEISGMSILIGKEFKLEL	678
RESULT 7			
ID	AAR42995		
XX	AAR42995 standard; Protein; 985 AA.		
AC	AAR42995;		
XX			
DT	16-MAY-1994 (first entry)		
XX			
DE	Glycosyltransferase.		
XX			
KW	Glycosyltransferase; GT; Gtase; Aspergillus niger; plasmid;		
KW	transformation; expression.		
XX			
OS	Aspergillus niger.		
XX			
PN	JF05268951-A.		
XX			
PD	19-OCT-1993.		
XX			
PF	27-MAR-1992; 92JP-0101760.		
XX			
XX			
PR	27-MAR-1992; 92JP-0101760.		
XX			
XX	(AMAN ) AMANO PHARM KK.		
PA	(UOZU/) UOZUMI T.		
XX			
DR	WPI; 1993-364282/46.		
DR	N-PSDB; AAQ50466.		
XX			
PT	Prepn. of glucosyl:transferase - using glucosyl:transferase gene		
PT	from recombinant DNA of Aspergillus niger		
XX			
XX	Disclosure; Page 6-9; 10pp; Japanese.		
PS			
XX			
CC	Glycosyltransferase was prepared from a culture of A. niger		
CC	containing recombinant DNA encoding Gtase (AAQ50466). Gtase can be		
CC	obtained effectively using this method and in larger quantities.		
CC	The probes (AAQ50466-67) were used to detect transformants		
CC	containing the Gtase gene.		
XX			
SQ	Sequence 985 AA;		
Query Match 30.5%; Score 1410.5; DB 14; Length 985;			
Best Local Similarity 37.3%; Pred. No. 9 6e-113;			
Matches 306; Conservative 139; Mismatches 243; Indels 133; Gaps 22			
QY	59 DVORLAVYASLETDSRLRVRI-----TDADHPRWE-VPODIIPRAPGQVLDHAPPASSAF 113		



Db 167 LG-----AHNDPMRLSVGVIRTFMNMQDSYGVNPGANLYGSHPVYIDHREITGT-HG 216  
Qy 229 VLLSSNGMDVLY-----GGSVYTYKVIUGVLDYFFAGPNPLAVDQYTQTLARAPMP 283  
Db 217 VLFNSNGMDVLIIDEDEGGKYLEYNTLGGVLDYFFVFGDSSPKAVEYGEIARPPMQP 276  
Qy 284 YWSEFGHCRKRYGLUNSDLERVARYAKARIPLEVMWTDIDYMDGFKDFTLDRVNFATAE 343  
Db 277 YWGLGFHCKGYQDAFWAEVYVNYGQAEIPELVMTDIDYMDRRVFTVDPDRPLPK 336  
Qy 344 LRPFVDRLHRNAQKYLILDPGIRVDPIDATYGTFFVGMQODIFLKR-NGTNFVGNVWPG 402  
Db 337 IRAWVDYLHEHDQYIIVKVDPAIAY-----VESGTLDRGLDDVFLRSNGSVMLGVVWPG 392  
Qy 403 DVYFPDMHPAAAFWAREISLFF---RRTIPVGLWIDMNEISNFYFPEPMNALDDPPYR 459  
Db 393 VTVEPDMFAENITQYMNNEFALFFDAEGVDIDGLWIDMNEPSNF---FCNFPDNPYE 448  
Qy 460 -----INNDGT-----GRP-----468  
Db 449 AAKGFPPTPPVPPRELPGFACVLOPEGTCEDEGTAGSKEDSGFGQGLVTRQPGF 508  
Qy 469 -----INNKTVRPLAVH 480  
Db 509 SRPRHPFHRROEYEGDOOKGLFGRDLLYPEYAIHNKAAFRDWNADKGGISNKTNTNVIH 568  
Qy 481 YGGVTEYEENHNLGLLEARATGRGV-LRDTGRRPFLVLSRSTFVSGRYTAVTGDNAATW 539  
Db 569 QNGLAEDYVHNLGAMSSASRDAMEARRPGLRPFITRSTTPHAGSKVGLWLDNLSNW 628  
Qy 540 GDLRYSTINTLSF-GLFGMPMIGADICGFNGNTTEELCGRMIOQLGAFYFPSRDSIAFTV 598  
Db 629 NOYRESIRTLAYTISIFQGMVGSDVCGFGDITNEELCARWASLGAFQTFERNHAQYEA 688  
Qy 599 RRELYLWPSVAASGRKALGLRYQLLPYPTLUMYEAHMTGAPIARPLPFSFPHDVATYGV 658  
Db 689 PQEYQWESVAESARRATGARYLLDYMYTALWKQSEQGTPAVVPVMPFVPEDEKGTLE 748  
Qy 659 RQFLGRGLVSPVLEPGPTTVDAYFAGRWYLYDY--SLAVATRCKHVLPAADTV 716  
Db 749 NOYFVGPGLVAPVVEQGSTVDVLPPEGVF--YDMWTHAIOGEGSYSVTGNTMI 806  
Qy 717 NVHLTGTITLPLQO-SALTTSRRARTAPHLVLAEDGTASGYLFLDDGDSPEYGRSDW 775  
Db 807 PLFIRGGVILPLRENSAMTTTELKKEFELLVLDNDGKAKGELYIDDGESLE---QESY 863  
Qy 776 SWRPNYKIPNNKGAIKVKSEVWNSYQSRTLVSKVLMGHRSPAAPKLTHVNSAE 835  
Db 864 TAVKFEYA---HGVVTLTDE-----FSEDWPVEVASVLL-----RPKGEIIV---E 905  
Qy 836 VEASSAGTRYQNAG 850  
Db 906 VKSFTAGRIKLG 920

RESULT 10

ABBS7174  
ID ABBS7174 standard; Protein; 953 AA.

AC ABBS7174;

DT 07-MAR-2002 (first entry)

DE Mouse ischaemic condition related protein sequence SEQ ID NO:422.

XX Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;

KW vasospastic ischaemia; ischaemic condition; ischaemic disease.

XX Mus musculus.

XX WO200108188-A2.

PN 22-NOV-2001.

XX

XX 18-MAY-2001; 2001WO-JP04192.  
PF 18-MAY-2000; 2000JP-0145977.  
XX (UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.  
XX Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;  
XX WPI; 2002-034733/04.  
DR N-PSDB; AB199468.  
XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring  
PT expression levels of particular genes defined in the specification or  
PT by determining the expression profile of a gene group comprising these  
PT genes -  
XX Claim 2; Page 1154-1158; 2690pp; English.  
PS The present invention describes a method for examining ischaemic  
XX conditions, comprising measuring the expression levels of particular  
CC genes (1) in a test sample or determining the expression profile of a  
CC gene group in the sample comprising genes selected from (1). The method  
CC is useful for examining the ischaemic condition (e.g. compressive  
CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring  
CC expression levels of particular genes (AB199202 to AB199912, encoding  
CC the protein sequences in ABBS7020 to ABBS7374) or by determining the  
CC expression profile of a gene group comprising these genes. The  
CC expression levels or expression profiles produced by these genes are  
CC used as an indicator when screening for ischaemic condition-improving  
CC drugs or therapeutics for ischaemic diseases. AB199913 and AB199914  
CC represent PCR primers for a mouse ischaemic condition related sequence,  
CC which are used in the exemplification of the present invention.  
XX SQ Sequence 953 AA;  
Query Match 30.3%; Score 1403; DB 23; Length 953;  
Best Local Similarity 38.1%; Pred. No. 4.1e-112;  
Matches 327; Conservative 122; Mismatches 323; Indels 86; Gaps 20;  
Qy 14 CLPAPRLCSKKEGGLAARTVLAATVMEALRAEAATGGRSSTGVDQVQLAVASLETDS 73  
Db 127 CFFPPSPSYRLN-----LSSTESGYTATLRTSPTFFPKDVLTLQLEVMETDS 177  
Qy 74 RLVRITTDADHRWEVQDIIPRAPGDVLDHAPASSAPLQGRVLSPAGSDLVLTWAS 133  
Db 178 RLHFKIDKPAKRIEVLLE-TPR-----VLSQAP-----SPLYSVEFS-----EE 216  
Qy 134 PRFTVSRSTGDTLPDT--APGLVFRDKYLEVTSALPAGRASLYGLGHTKSSFLRHN 191  
Db 217 PFGVIVRRKLGKRVLLNTTVP-LFFADQLQLSTSLPA--QHITGLGHLSPMLSTDW 273  
Qy 192 DSFTLWADIGASYVDVNLGSHPPFVMDVRACTAGHVLLLSSNGMDVLYGGS-VYTYKV 250  
Db 274 ARITLWNRDTPPSQ-GTNLYGSHPFYLALEDGLAHGVFLNSNAMDVTLQSPALTWRS 332  
Qy 251 IGGVLDYFFFGAGNPLAVVDQYTLIARPAWPPYSGFHQCRKYGLYLVNSDLERVARYA 310  
Db 333 TGGILDVYFLGPEKSVVQYLDVVGYPMPYWGGLGFLHCRWGYSSTAIVROVVENMT 392  
Qy 311 KARIPLEVMWTDIDYMDGFKDFTLDRVNFATAELRPFVDRLHRNAQKYLILDPGIRVD 370  
Db 393 RTHFFLDVQWMDLDYMDARRDFTFNQDSF--ADFPDMVRDVHQGGRRYVMI VDPAISSAG 450  
Qy 371 IDATYGTFRGMOODIFL-KRNGTNFVGNVWPGDVYFPDMHPAAAFWAREISLFRRTI 429  
Db 451 PAGESYRYPDEGLRGVFIITNETGQPLIGKVCPTTAFPOFTNPETLDWQDMVSEFHQV 510  
Qy 430 PVDGLWIDMNEISNFYNPE----PMNALDDPPYRINNDDGTGRPINNKTVRPLAVHYGGVT 485  
Db 511 PFDGMMLDWNEPSNFVRGSGQCGCPNNELENPPY---XPGVWVGILQAATICASSHOFST 567  
Qy 486 EYEBHNLFGLLEARATGRGVLRDTRRRPFLVLSRSTFVSGRYTAVTGDNAATWGLDRYS 545

Db 568 HYNLHNLGLTEAASSRALVKTGRTPFVIGSRSTFSGHGRVAGHMTEDVRTSWEHLAYS 627  
 QY 546 INTMLSFGLGEMPMICADICGFNCTTEELCGRWIOLGAFYPSRDHSAIFVRRLELYLW 605  
 Db 628 VPDIOFLNLGLVPLVGADICGFIGDTSSELCLVRWTLQAGAFYPMRHNOLNSVPOEYRF 687  
 QY 606 PSVAASG-RKALGLRQLLPFYFTLMEAHMTGAPIARPLFTSYPHDVATYGVDRQFLLG 664  
 Db 688 SETAQAMKAFALRVALPYLTLEFRAHVRGDTVARPLFLEFREDPSTVSDROLLWG 747  
 QY 665 RGVLSVPLEPGPTTVDAYFPAGRWLYLDSL-----AVATRGTGRHV 707  
 Db 748 PALLITPVLPEPGKTEVTGYFPKGTWYMNQVSVDSLTLPSPSSASSFRSAVQSK-GQWL 806  
 QY 708 RLPAPADTUNVHLTGTLTLPLOQSALTTSRARTAFHLLVALAEDGTAGSYLFLDDGSP 767  
 Db 807 TLEAPLDTINVHLREGYIIPLOQPSLTTTTSRKQPMALALTASGEADGELFMDGDES 866  
 QY 768 EYGRSDMSVRNRYKIPNN-----KGAIKVKGSEVHNSVAQKRTLVISKVLIM 816  
 Db 867 AVLEHAYTLVFSAK--NNTIVNKLVRVTKGAEQLQKEVTVLGVATPTQVLSNGIPV 924  
 QY 817 G-----HRSPAAPKKL 827  
 Db 925 SNFTYSPDNKSLAIPVSL 942

## RESULT 11

AAW15191  
 ID AAW15191 standard; Protein: 985 AA.

AC AAW15191;

DT 18-JUN-1997 (first entry)

DE Aspergillus oryzae alpha-glucosidase.

KW Fungus; fungal; enhancer element; promoter; recombinant protein.

OS Aspergillus oryzae.

PN JP09009968-A.

PD 14-JAN-1997.

PF 29-JUN-1995; 95JP-0163579.

PR 29-JUN-1995; 95JP-0163579.

XX (KCKU-) KOKUZEI CHO CHOHAN.

PA (OZEK-) OZEKI KK.

XX WPI; 1997-126425/12.

DR N-PSDB; AAT65017.

XX Fungal DNA enhancer element - used to transform other host fungus  
 PT e.g. *Aspergillus oryzae*, to produce large quantities of a gene  
 PT product

XX Disclosure; Page 14-19; 25pp; Japanese.

XX The promoter region of the alpha-glucosidase (agda) gene of  
 CC *Aspergillus oryzae* contains two novel enhancer elements: one  
 CC (designated "enhancer-B") corresponds to the consensus sequence  
 CC CGGNATTATA and the other (designated "enhancer-C") is of sequence  
 CC CCAATCAGCGT. By inserting at least one of the enhancer elements  
 CC into a promoter region which is functional in fungi, the activity  
 CC of the promoter is enhanced. Using such improved promoters, a gene  
 CC of interest can be expressed efficiently in transformed fungi.  
 CC The present sequence is encoded by the agda gene.

XX Sequence 985 AA;

ID AAW74090 standard; Protein: 1827 AA.

Query Match 29.1%; Score 1348.5; DB 18; Length 985;  
 Best Local Similarity 34.1%; Pred. No. 2.3e-107;  
 Matches 312; Conservative 158; Mismatches 289; Indels 157; Gaps 28;

QY 59 DVQRLAVYASLETDSRLVRVIT----DADHPRWEV-PQDIIPREPAGVDLHADPASPASAP 113  
 Db 97 DVDSLTLTVEYOAKDLNLIQIVPTYPDASNASWYILSEELVPRP-----KASONASVP 149  
 QY 114 LQGRVLPSPAGSDLVLTWHASP-FRFTVSRSTGTDLFDT-APGLVFRDKYLEVTSALPAG 171  
 Db 150 -----QSDFYVWSNPEFNFVKIRKATGDLVFNKSTGLVYENQFIEFTVLLPE- 199  
 QY 172 RASLYGLGEHTKSSFRRLRHNDSTFLWNADIGASVVDVNLVYGHSPFYMDVR----- 221  
 Db 200 EYNLYGLGER-MNQRLLENANLTYAADI-ADPIDNIYGHAFYLDTRYKYKGGQNK 257  
 QY 222 -----ARGT-----AHGVLLSSNGMDVLYGGSYVTVYKVIIGGVLDFFYFAGNPPLA 267  
 Db 258 HTIVKSSSEAPSQEVVSYSHGVFLRNAHQEILLRDOQKLIWRTLGGSVDLTFYSGPTQAE 317  
 QY 268 VVDQY-TQLIARAPAMPYKSEFGHOCRYGLNVSDLERVVARYAKARIPLEVMWTDIDYM 326  
 Db 318 VTKOYQLSTVGLPAMQOYNTLGFHOCRWYNNWSEFEDVLANPERFEIPLBYLWADIDYM 377  
 QY 327 DGFKDFTLDRVNFATAELRPFVDRLHNAOKYVLILDPGIRV-DPIDAT--YGTFFVRGMQ 383  
 Db 378 HGYRNFNDQHRFSYEGEKFLNKLHAGRRWPIVDGALYIPNPNASDAYEYIDRGAK 437  
 QY 384 QDIFLKR-NGTNFVGNVWPGDVYFPDFPMHPAAAFWAEISLFRRTIPVDGLWIDMNEIS 442  
 Db 438 DDVFINKPDGSLYIGAVWPGVTVYPDWHHPKASDFWANELVTWNKLHYDGVWYDMAEVS 497  
 QY 443 NF-----YNP-----EPMNA----- 452  
 Db 498 SFCVSGCOTGNLSMNPAPHPPALPGEQGNVVYDYPEGFININATEAASASAGAAASQSA 557  
 QY 453 -----LDDPPYRINNDGTGRPINNKTVRPLAVHYGGVTEYEEHNL 492  
 Db 558 SSTTTSAPYLRTTTPGVRNVDPHPYVNHVQPGDLSVHAISPNSTHSDGVQVYDVHSL 617  
 QY 493 FGLLEARATGRGVLRD-TGRPPFVLSRSTFVGSGRYTAIYWTGDNAATWGDRLRYINTMLS 551  
 Db 618 YHQGINATYHGLLKVENKRPFIARSTFSGSGKWAGHMGDNFSGKWSMFFSISQALQ 677  
 QY 552 FGLFCMPMIGADICGFNCNTTEELCGRWIOLGAFYPSRDHSAIFTVRRRELYLWPSVAAS 611  
 Db 678 FSLFGIPMGVDTGCFNGNTDEELCNRMWQLSAFPPFYRNHNVLISAIPQEPYRWASVIDA 737  
 QY 612 GRKALGLRYQLLPFYFTLWYEAHMTGAPIARPLFFSYPHDVATYGVDRQFLLGRLVSP 671  
 Db 738 TKAAWNIIRYAILPYFTLPHLAHTTGSTVMRALAWEPFNDPSLAAVGTQFLVGFSPVMVIP 797  
 QY 672 VLEPGFTTVDAYFPA---GR-WYRLYDYSLAVATRTKXHLVLPAPADTVNHLTGTTLP 727  
 Db 798 VLEPQDVTQVQVFGVGHGEVMDWYSOT-AVDAKPGVNTTISAPLGHIPVIFVRGGSILP 856  
 QY 728 LQOSALTTSRARTAFHLLVALAEDGTAGSYLFLDDGDS--PEYGRSDMSWVRFNKIP 785  
 Db 857 MQEVALITTRDARKTPWSLLASLSNGTASGQYLDGDSVTPEDTSLVDFLASRSLR-A 915  
 QY 786 NNGKAIVKSEVHNSYAQSRTLVIKSVLVMGHRSPAAPKKLTIVHNSAEVSEASSAGTR 845  
 Db 916 SARGTWKEANP-----LANVTVLG----VTEKPSSTVLNG---ETLSSDSVK 955  
 QY 846 YONAGLGGVVAHIGSL 861  
 Db 956 Y---NATSHVLHVGL 968

## RESULT 12

AAW74090











GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 ; Search time 62.5261 Seconds  
(without alignments)  
3772.184 Million cell updates/sec

Title: US-10-043-418-2

Perfect score: 4862

Sequence: 1 MERSKLPYICPTLAVLPL.....RLVQQAFLKLEFEGATRV 914

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL 23.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2719.5	55.9	902	10 Q9LYF8	Q9lyf8 arabidopsis
2	2714.5	55.8	902	10 Q22444	Q22444 arabidopsis
3	2485	51.1	879	10 Q9LYY2	Q9lyy2 hordeum vul
4	2191.5	45.1	910	10 Q8VWV9	Q8vwv9 pinus pinas
5	2058.5	42.3	929	10 Q9LGC6	Q9lgc6 oryza sativ
6	2050.5	42.2	916	10 Q9AVC3	Q9avc3 physcomitre
7	2045	42.1	907	10 Q92P26	Q92p26 arabidopsis
8	2045	42.1	907	10 Q92P26	Q92p26 arabidopsis
9	2000	41.1	928	10 Q957Y7	Q957y7 arabidopsis
10	1969.5	40.5	935	10 Q9LEC9	Q9lec9 solanum tub
11	1849	38.0	855	10 Q92P04	Q92p04 tropaeolum
12	1553.5	32.0	932	10 Q9LZT7	Q9lzt7 arabidopsis
13	1485	30.5	953	11 Q8BGI6	Q8bgi6 coturnix co
14	1484	30.5	953	11 Q91Z45	Q91z45 mus musculu
15	1454.5	29.9	937	6 Q9MYM4	Q9mym4 bos taurus
16	1451.5	29.9	952	4 Q81WE7	Q81we7 homo sapien

17	1373.5	28.2	873	13 Q73632	Q73632 coturnix co
18	1369	28.2	992	3 Q9UV08	Q9uv08 emericella
19	1364.5	28.1	1734	4 Q8TE24	Q8te24 homo sapien
20	1329	27.3	995	3 Q9URX4	Q9urx4 schizosacch
21	1260	25.9	955	3 Q9CL57	Q9cl57 emericella
22	1138	23.4	1743	5 Q19004	Q19004 caenorhabdi
23	1112	22.9	955	5 Q21750	Q21750 caenorhabdi
24	1105.5	22.7	920	5 Q9NFY8	Q9nfy8 penaeus van
25	938	19.3	856	5 Q20722	Q20722 caenorhabdi
26	892.5	18.2	763	16 Q8VAE7	Q8vae7 listeria mo
27	868.5	17.9	763	16 Q92F84	Q92f84 listeria in
28	864	17.8	728	2 Q9RH22	Q9rh22 alicyclobac
29	862	17.7	921	10 Q9FN05	Q9fn05 arabidopsis
30	852.5	17.5	914	4 Q81ZM5	Q81zm5 homo sapien
31	852.5	17.5	914	4 Q81ZM4	Q81zm4 homo sapien
32	849.5	17.5	769	4 Q81WZ0	Q81wz0 homo sapien
33	849.5	17.5	925	4 Q8TE14	Q8te14 homo sapien
34	828.5	17.0	779	16 Q8YN00	Q8yn00 anabaena sp
35	828	17.0	966	11 Q08794	Q08794 mus musculu
36	827	17.0	991	10 Q93Y12	Q93y12 arabidopsis
37	824.5	17.0	746	16 Q8XIN9	Q8xin9 clostridium
38	823	16.9	944	4 Q14697	Q14697 homo sapien
39	821	16.9	966	4 Q9P0X0	Q9p0x0 homo sapien
40	820	16.9	944	11 Q8BHN3	Q8bhn3 mus musculu
41	820	16.9	991	3 Q8NIY3	Q8niy3 neurospora
42	818	16.8	944	6 P79403	P79403 sus scrofa
43	816	16.8	919	10 Q24375	Q24375 solanum tub
44	815.5	16.8	751	16 Q8RDL1	Q8rdl1 thermoaer
45	811.5	16.7	653	6 Q9BE70	Q9be70 macaca fasc

#### ALIGNMENTS

#### RESULT 1

Q9LYF8 PRELIMINARY; PRT; 902 AA.  
 AC Q9LYF8  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)  
 DE Alpha-glucosidase 1 (AT5G11720/T22P22\_110).  
 GN T22P22\_110.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;  
 OC eurosids\_II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_taxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,  
 RA Bancroft I., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.;  
 RA Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 RL [2]  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Shinn P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,  
 RA Banh J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D.,  
 RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,  
 RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,  
 RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,  
 RA Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,  
 RA Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,  
 RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,  
 RA Chan M.M., Chang C.H., Dale J.M., Hayashizaki Y., Hsuan V.W.,

QY	12	PTLAVVLVPLVCMVVEGATTSKYNDQGBAIGYGYOVKNAKVNDSTGKSLTALLQILRNSP	71
DB	8	PNXIFV--VVVFESLRSQVVLSEESFTVWGYGVYVRSVGD--SNRQVITAKLDLKPSS	64
QY	72	VYGPDIHFUSFTASPEEDDTRLRFTDANRRWEIPNEVLRPRPPPPSPPLSSLOHLPK	131
DB	65	VYAPDI:KSLNLHVSLTSELRIRITDSSQORWEIPETVIPR-----AGNHSR	113
QY	132	PIPQ-----NQPTTVLGHPSDDLAFTLFHTTPEGFTYKRKSTHDLVLFDAIPSPNPTTF	186
DB	114	RSTEEEDGGNSPENNFLADPSSDLVFTLHNTTPFGFSVSRSSGDLFDTSPSSDSNTY	173
QY	187	LYKQOYLQSSLPQAQAHLYGLGEHTKPTFOLAHNQILTLWNADIASFRNDLNLGSH	246
DB	174	FIFKQOFLQSSALPENKSNLYGGEHTKFRFLIPGETMTLWNADIGSENPOVNLGSH	233
QY	247	PFYMDVRSS--PMYGSTHGVFLNSNGMDVYETGDRITKYKGGIIDLIFYAGRTPEMV	303
DB	234	PFYMDVRGSKGNEEAGTTHGVLLNSNGMDVKYEGHRIYNYIGGVLDLVVPAGSPENV	293
QY	304	LQOYTKLIGRPAMPYWAFGHQCRWGYRVNWEIETVVDKYABARIPLEVMTDIDYMDA	363
DB	294	MNQYTELIGRPAMPYWFGFHQCRGYGVKNSDLEYVVDGYKAGIPLEVMTDIDYMDG	353
QY	364	KXDFTLDPVHFPLDKQOQFVTKLHRNGORYPILDPGINTKSKSVGTFRGQSNVFIKEN	423
DB	354	YKDFTLDPVNFPEDKXQGFVDTLHKNGQKQYVLLIDPGIWDSSYGYINRGNEADVFIKEN	413
QY	424	GNPYLGSVMPGVYYPDFELDPAARSFWDEIKRFRDILPIDGIMDNNEASNFITSAPTF	483
DB	414	GEPLYGEVMPGVKYYPDFLNPAAATFWSNEIKMQEILPLDGLWIDMNLSENFITSPSS	473
QY	484	GSTLDNPPYKINNSGRVPINSKTLPATAMHYGNVTEYNAHNLYGFLESQATREALVRPA	543
DB	474	GSSLDPPYKINNSGDKRPINNKTVPATISYHFGNISEYDAHNLYGLEAKATHQAQVVDIT	533
QY	544	TRGPFLLSRSTFAGSGKTAHWTDGNAARWDDLOVSIPTMLNFGLPQMPMIGADICGFAE	603
DB	534	GKXPPFILSRSTFVSGSKYTAHWTDGNAAKWEDLAYSIPGILNFGLPFIPMWGADICGFHS	593
QY	604	STTEELCCRMIQLGAFYPSRDSHARDTHOELYLWESVAASARTVLGLRYELLYPYTL	663
DB	594	DTTEELCCRMIQLGAFYPPFARDHSSLSGTARQELYLWDSVASSARKVJGLRMRLPLHYTL	653
QY	664	MYDANLRGSPiARPLSFTFPDDOVATYGTSSOFLIGRIGMIVSPVLQPGSSIVNAYSPRGNW	723
DB	654	MYEAHVSGNPIARPLFPFPQDTKTYTEIDSOFLIGKIMVSPALKQGAVAVDAYFPAGNW	713
QY	724	VSLSNYTSVSVSAGTYVYLSLAPPDHNVHIEGNIVAMOGAEMTTOAARSTPFHLLVM	783
DB	714	FDLFNTYSFVAGDGSKGHVRLDTPADHVNHVREGSIVAMOGAELTTDRAKRTPYQLLVYA	773
QY	784	SDHVASTGELFLDNGIEMDGGPGGK--WTUVRFPFASGGINLNTISSEVVNRYGMSORW	841

```
QY 424 GNPYLGSVMPGPYYYPDFLDPAARSFWDEIKRFRDILPIDGDIWIDMNEASFIITSAPT 483
D 414 GEYLGEVWPGKYYPDFLNPAAATFMSNEIKMFQELPLDGLWIDMNEISFIITSPSS 473
QY 484 GSTLDNPPYKINNNGRVPINSKTIIPATAMHYGNVTYNAHNLNGYLGESQATREALVRPA 543
D 474 GSSLDDPPYKINNNGRVPINNTKVPATSIHFGNISEYDAHNLNGYLGLEAKATHQAVVDIT 533
QY 544 TRGPILLRSSTFAGSGKYTAHWTGDNAAWDDLOYSIPTMLNPLGFCMPMIGADICGFAE 603
D 534 GRPFFILRSSTFVSSGKYTAHWTGDNAAWDDLOYSIPTMLNPLGFCMPMIGADICGFSH 593
QY 604 SITTEELCRRWQLGAFYPPSRDHSARDTHQOELYLWESVAASARTVLGLRYELLPPYYTL 663
D 594 DTTEELCRRWQLGAFYPPFARDHSSLSGTARQELYLWDSVASSARKVLGLRMLLPHLYTL 653
QY 664 MYDANLRGSPDIARPLSFTFPDDVATYGISQFYLGRGIMVSPVLQGSISVNAYSRGNW 723
D 654 MYEARYSGNPIARPLSFTFPDDVATYGISQFYLGRGIMVSPVLQGSISVNAYSRGNW 713
QY 724 VLSLNTSSVSAGTYVLSAPPDINHVIHEGNIVAMQGEAMTTQAAARSTPFHLLVVM 783
D 714 FDLNTPSPVGGSGKHRLDTFADHVNHVREGSIVAMQGEALTTDRDKTPYQLLVA 773
QY 784 SDHVASTGELFDNGIEMDGGPGGK--WTLVRFFAESGINNLTISSEVVNRGYAMSQRW 841
D 774 SRLNLSIGELFDNGENLRWAGGGRNDWTLVKFCYVTKSVVLRSEVVNPYASKMKW 833
QY 842 VMDXITILGLKRRVKIKEYTVQKDGAIKVKGGRRTSSHNOGGFFVSV- ISDLRLVQ 900
D 834 SIGKTVFVGFENVENKTVERTSERLSPRISLIKTVDNDOPRPLSVESKLSLLVGK 893
QY 901 AFKLEL 906
D 894 KFERL 899

RESULT 3
Q9LLY2 PRELIMINARY; PRT; 879 AA.
ID Q9LLY2 PRELIMINARY; PRT; 879 AA.
AC Q9LLY2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE High pI alpha-glucosidase.
GN AGL97.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Igr1;
RX MEDLINE=20267959; PubMed=10806244;
RA Frandsen T.P., Lok F., Mirgorodskaya E., Roepstorff P., Svensson B.;
RT "Purification, enzymatic characterization, and nucleotide sequence of
a high-isoelectric-point alpha-glucosidase from barley malt.";
RL Plant Physiol. 123:275-286(2000).
DR EMBL; AF118226; AAF76254.1; -
DR InterPro; IPR000322; Glyco_hydro_31.
DR InterPro; IPR002052; N6_Mtase.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
DR PROSITE; PS00092; N6_MTASE; 1.
SQ SEQUENCE 879 AA; 96558 MW; AF9235ECE7D15B44 CRC64;

Query Match 51.1%; Score 2485; DB 10; Length 879;
Best Local Similarity 53.6%; Pred. No. 2.4e-180;
Matches 480; Conservative 149; Mismatches 235; Indels 32; Gaps 13;
```

```
QY 16 VVPLVLVCMVVEGATTSKNDNGEAGIYGIOVQNAKVDNSTGKSLTALLQIRNSPVYGP 75
D 7 LLLCLCLCLFAPRLCSSK--EEGPLAAGGYRSVAVDDG--GRLRABAAAATGCASSTG 63
QY 76 DIHFLSPTASFEEDDTLRIPDTDANNRRWEIPNEVLPRPPPPP---SPPLSSLOHLKP 132
D 64 DVQRLAVYASLETDSRLVRITDADHPREWPQDIIIPRAPADVLHDAPPASS-----AP 118
QY 133 IPOQNTPTVLSSHPSDLAFTLPHITTPGFTYKSKTHDVLFDATPSPNPTTLLIYKQ 192
D 119 LQGRK-----VLSAAGSDLVLTV--HASPFRFTVSRRTSDIILDTAP-----GLVFRDK 166
QY 193 YLQSSSLPAQAHLGLGEHTKPTFQLAHNOIITLWNADIASFNRDNLNGSPHFYMDV 252
D 167 YLEVTSLAPGRASLYGLGEHTKSSFLRHNDSTFLWNADIGASVVDVNLNGSPHFYMDV 226
QY 253 RSSPMVSGTHGVFLNLSNGMDVEYTGDRITVKYVGGIIDLHIFAGRTPEMVLDOYTKLIG 312
D 227 RPP---GTAHGVLSSNGMDVLYGGSYVTVYKVGGLDFYFFAGPNPLAVVDQYTOIG 283
QY 313 RPAPMYWAFGPHOCRMGYRDVNEIETVVDKYAZARIPLEVVMWTDIDYMDAPKDFTLDPV 372
D 284 RPAPMYWAFGPHOCRMGYRDVNEIETVVDKYAZARIPLEVVMWTDIDYMDAPKDFTLDPV 343
QY 373 HFLDKMQQFVTKLHRNGQRYVPIIDPGINTNKSYGTFIRGMQSNVFIKRNGNYPYLSGW 432
D 344 NFTAELRPFVDRLHRNAQKYVILDPGIRIDATYGTIVRGMQDIFLKRNGTNFVGNVW 403
QY 433 PGVYVYDFDLPAPARSFWDEIKRFRDILPIDGDIWIDMNEASFIITSAPTGSTLDNPPY 492
D 404 PGVYVYDFDLPAPARSFWDEIKRFRDILPIDGDIWIDMNEASFIITSAPTGSTLDNPPY 461
QY 493 KINSGGRVPIINSKTIPATAMHYGNVTYNAHNLNGYLGESQATREALVRPATRGPPELLSR 552
D 462 RINDGTGRPINNKTVPASVHYGCVTEYDAHNLFGLEARATHRALRDTGRPPFVLRSR 521
QY 553 STFAGSGKYTAHWTGDNAAWDDLOYSIPTMLNPLGFCMPMIGADICGFAESITTEELCCR 612
D 522 STFAGSGKYTAHWTGDNAAWDDLOYSIPTMLNPLGFCMPMIGADICGFAESITTEELCCR 581
QY 613 WQLGAFYPPSRDHSARDTHQOELYLWESVAASARTVLGLRYELLPPYYTILMYDANLRGS 672
D 582 WQLGAFYPPSRDHSARDTHQOELYLWESVAASARTVLGLRYELLPPYYTILMYDANLRGS 641
QY 673 PIARPLSFTFPDDVATYGISQFYLGRGIMVSPVLQGSISVNAYSRGNVWSLSNYS 732
D 642 PIARPLSFTFPDDVATYGISQFYLGRGIMVSPVLQGSISVNAYSRGNVWSLSNYS 701
QY 733 VVSAGTYVLSAPPDINHVIHEGNIVAMQGEAMTTQAAARSTPFHLLVMSDHRVASTGE 792
D 702 VATRTGKHVTLPAADTVNVHVAGGTILPLQSSALTTSRARRTAFHLIVALAEDGTASGD 761
QY 793 LFLONGIEMDGGPGGKWTLVRFFAESGINN--LTISSEVVNRGYAMSQRWMDKITLGL 851
D 762 LFLONGIEMDGGPGGKWTLVRFFAESGINN--LTISSEVVNRGYAMSQRWMDKITLGL 820
QY 852 KRRVKIKEYTVQKDGAIKVKGGRRTSSHNOGGFF--FVSVISDLRQLVGOAFKLEL 906
D 821 RSPAPKLTLVHNSAEVEASS--SAGTFYQAGGLGGVHAIGGLSLVVGEEFELKV 875

RESULT 4
Q8VMV9 PRELIMINARY; PRT; 910 AA.
ID Q8VMV9 PRELIMINARY; PRT; 910 AA.
AC Q8VMV9;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative alpha-xylosidase.
GN XYL1.
OS Pinus pinaster (Maritime pine).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
```



```
QY 437 YPDLFDPAARFVWDEIKRFRDILPIDIGIWMNEASNFITS-----APTFGSTL 487
D 416 YPDLFLNPNVSWWIDVRRFDHLPVDPDGLWIDMNEASNFKGKCEIPTHLCPNPNTT 475
QY 488 -----DNPPYKINNSGRVPINSKTIIPATAMHYGNVTEYNAHNLVGFLE 531
D 476 PWCCLDCKNLNTRWDEPPYKINASGQTARLGFNTIATSATHYNGILEYNAHSLYGFQ 535
QY 532 SOATREALVRPATRGPFLLSRSTFAGSGKYTAHTWGDNAARDLQYSTPTMLNFGLEGM 591
D 536 AIATHALQLOGLQKRFILTRSTFVSGYAAHAWTGDNKGKTWENLURYSISTVNLFGIFGM 595
QY 592 PMIGADICGFAESTTEBELCCRWELQAGAFYFGRDHSARDTHQELYLWESVAASARTVLG 651
D 596 PMVGADICGFPPTPEELCNRWELQAGAFYFGRDHSARDTHQELYLWESVAASARNALG 655
QY 652 LRYELLPPYTYTLMYDANLGRSFIARPLSTFTPDVATYGISQFLIGRIMVSPVLQPGS 711
D 656 MRYRLLPYLYTLNYQAHLTGAPVAPVFFSFDFTFCYGLSTQYLLGASVMVSPVLEQGA 715
QY 712 STVNAYSPPGNMVSLSNYTSSVSASGT-YVLSLAPPDHIHVHIEGNIAMOGAEMTQ 770
D 716 TSUSAMPPGSGWNLFD-TTKVVSRSRGAVKLDAPLNEINHVHVFQNTLLPMQGGTISK 774
QY 771 AARSTPFLHLLVM---SDHVASTGELFLDNGIEMDIGGFGKWTLVRFPAESGINNLITIS 827
D 775 EARATPFLVAVFPFCATEAEAGAVVDDDERPEMVLAEQATVRFVATVGRKAVTVR 834
QY 828 SEVNRGYAMQORVWMDKTIILGLKRRVKIKEYTVQKDAKAIVKUGLRRTSSHQGGFF 887
D 835 SEVELGSYSYLGKLLIEKLSVLGL-----EGTGRDLAVHVDGANA 874
QY 888 VSVISDLRLQVGAQFKELEFEF 910
D 875 TAIATSRPYFAG-----AEAEHLG 893
RESULT 6
Q9AVC3 PRELIMINARY; PRT; 916 AA.
ID Q9AVC3;
AC Q9AVC3;
CT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE Putative alpha-glucosidase.
GN pGLU.
OS Physcomitrella patens subsp. patens.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
OX Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrellia.
RX NCBI_TaxId=145481;
RN [1]
RP SEQUENCE FROM N.A.
RA Hiwatashi Y., Nishiyama T., Hasebe M.;
RT "Establishment of gene- and enhancer-trap systems of the moss,
RL Physcomitrella patens.,"
DR EMBL; AB057452; BAB39467.1; -
DR InterPro; IPR001064; Crystallin.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
SQ SEQUENCE 916 AA; 102282 MW; AAEE2AC6E440D883 CRC64;
```

```
Query Match 42.2%; Score 2050.5; DB 10; Length 916;
Best Local Similarity 44.9%; Pred. No. 3.3e-147;
Matches 408; Conservative 163; Mismatches 273; Indels 65; Gaps 17;
QY 13 TLAVPLVLCHVVEGATTSKNDNQGEALGYGVQVNAKVDNSTGKSLTALLQLRNSPV 72
D 4 TIIFAMALLCFL--NLTAALSPKALRSFGAGHRMTSV-FEPDGGSGFVANLELITGTET 60
```

```
QY 73 YGPDHIFLSTASFEEDTLIRFTDANRRWEIPNEVLPRPPPPSPPLSSLQHLKPK 132
D 61 YGPDISPLMIARYSDDLRHVHTDSIHARWEVPQDIIIPROSS-----SLVTHVKS- 113
QY 133 IPQNPQTTVLSPHPSDLAFTLHHT--PGFTYRKSTHVDLFDATP---IPSNPTTF-- 186
D 114 --RDMHSEGLDPAANDRQLQLSYTVFEPFGFAITRTSTGECLFNSTPIQDSGEPAFNS 171
QY 187 LIYKDOYLOLSSSLPAQAAHLYLGEHTKPT-FQLAHNQILTLWNADIASFNRLNLVGS 245
D 172 MVFKDQLEISTQLPRNNS-LFGIGBSTRPDGLRTRGLTYTLWATIDAAKYKVDVDLYGA 230
QY 246 HPFYMDVRSFPMVSGTHGVFLNSNGMDVEYTGRIYKYVIGGIIDLYIFAGRTPEMVL 305
D 231 YPFYMDIREG---GATHGVLMNSNGMDIIVWGEDMLTYHYGVLDYFFAGPAPLAVID 287
QY 306 QYTKLIGRPAMPYWAFFGHQCRMGYRDVNEIETVDKYAEARIPLEVMMDTIDYMDAFK 365
D 288 QYTNLIGRTPMPYWSFGFHQCRMGYETIDEIKDVAANYKKANIPDLTIWNDIDYMDAYK 347
QY 366 DFTLDPVHFPDLKMQOQVTKLHRNGORYVILDPGINTN-KSYGTFIRGMSNVFIKRN- 423
D 348 DFTPDPPYDENTVREFVKELHANGQYIIVLDPGISGVYKNYSTLERGLKDDIFLKNKF 407
QY 424 GNPYLGSVMPGPVYYPDFLPAARSFWVDEIKRPRDILPIDIGIWMNEASNFITSAPT 483
D 408 GNNYLAQVWFGVYFPDFLHPKASSWWTQELADFFDKVPFDGLWIDMNEASNPCTGSACS 467
QY 484 GSTL-----DNPPYKINNSGRVPINSKTIIPATAMHTGNV 518
D 468 FDTLTLMGKNDSNDRCLLHCNVNGTSRFDPPYKINHVGTYDNLGVKTIAMTVKHYNGV 527
QY 519 TEYNAHNLVPLESOATREALVRPATRGPFLLSRSTFAGSGKYTAHTWGDNAARDLQY 578
D 528 LEYDAHNLVGLCESIATQKTLRDVTGKRFILSRSTFVSGAHTAHWIGDKATWEDLKY 587
QY 579 SIPTMLNFGLFMPMIGADICGFAESTTEELCCRWELQAGAFYFGRDHSARDTHQELYL 638
D 588 SIVSVINSGMFVPMVGADICGFAAGNTEELCCRWELQAGAFYFGRDHSARDTHQELYL 647
QY 639 WESVAASARTVLGLRYELLPPYTYTLMYDANLGRSFIARPLSTFTPDVATYGISQFLIG 698
D 648 WESVAEASRRKALGLRYLLPHLYTLMEATKSGAPIARALFFSPKDLNLTALNDQPLL 707
QY 699 RGIMVSPVLQPGSSIVNAYSPPGNMVSLSNYTSSVSASGT-YVLSLAPPDHIHVHIEG 758
D 708 RSVLSIFVIAEGLTSVNAYFPKGTWNLDFD--SKIVSTGERRMPLAPADINSIHVSEGO 765
QY 759 IVAMQGEAMTTQAAARSTPFHLLVMS-DHVAS-TGELFLDNGIEMDIGGPGKWTLVYRF 816
D 766 ILPMQEARLTSAEVKKTPFTLVVVSADASASAGKLFVDSGVDIEMGIQDGSSTFVQFF 825
QY 817 AESGINNLTSSEVVNRGYAMQORVWMDKTIILGL-----KERVKIKEYTVQKD 865
D 826 AERSLHSGSLVSRVIAGNYALEQGLVQLSIRFLVSGVSGVSDIVNGERIVSAEQLSYDAR 885
QY 866 AGAIKVKGL 874
D 886 LESIQVSG 894
RESULT 7
Q9ZP26 PRELIMINARY; PRT; 907 AA.
ID Q9ZP26;
AC Q9ZP26;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Alpha-xylosidase precursor (Fragment).
GN XYL1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
```

OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.	
OX	NCBI_TaxID=3702;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=cv. Columbia; TISSUE=Seedling hypocotyl;	
RX	MEDLINE=21295583; PubMed=11402218;	
RA	Sampedro J., Sieiro C., Revilla G., Gonzalez-Villa T., Zarra I.;	
RT	"Cloning and Expression Pattern of a Gene Encoding an alpha-Xylosidase	
RI	Active against Xyloglucan Oligosaccharides from Arabidopsis."	
RL	Plant Physiol. 126:910-920(2001).	
DR	EMBL; AF087483; AAD05539.1;	
DR	InterPro; IPR000322; Glyco hydro 31.	
DR	Pfam; PF01085; Glyco hydro 31.1	
DR	PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.	
KW	Signal.	
FT	NON_TER	1
FT	SIGNAL	<1 19
FT	CHAIN	116 907
SQ	SEQUENCE	907 AA; 101647 MW; 58ABBD235366C588 CRC64;
Query Match	42.18; Score 2045; DB 10; Length 907;	
Best Local Similarity	43.81; Pred. No. 8.7e-147;	
Matches 409; Conservative 166; Mismatches 282; Indels 76; Gaps 16;		
Qy	13 TLAVLPLVLCMVVEGATTCKNDNQGEAIGYGVQVNAKVDNSTGKSLTALLQLIRNSPV 72	
Db	3 SLSLLALLLCF---SPTQSY-----KTIQGYRL--VSTESPDDGGFYGVLQVQKNKI 52	
Qy	73 YGPDHFLSTASPEEDDTLRIRFTDANNRWEIPNEVLPRPPPPSPPLSLSLQHLKP 132	
Db	53 YGSDITTLRLFLVRKHETDSRLRHVHTDAKQWEPVNLPRQPPQVQGVIGKSRKSP-- 110	
Qy	133 IPQKQPTTVLSSHPSDLAFLHTTTPFGFTIRKSTHVDLFDATPIPSNPTFLIYKQ 192	
Db	111 -----ITVEISGSELIFS-YTTDFTFAVRSNHEITL-----NTTSLVFKQ 155	
Qy	193 YLQ-SSSLPAQAAHLGLGHEHTKPT-PQLAHNQILLTNADIASFNRLDMLYSGHPFYMD 251	
Db	156 YLEISTSLP-KEASLYLGIGNSQANGIKLVNPEYTLTDEVSAINLNTLYSGHPFYMD 214	
Qy	252 VRSPMVGSTHGVLNLSNGMDVETGDRITYKIGVIGIILYIPAGTTPMVLQDQTKLI 311	
Db	215 LRNVGGKAYAHAVLLNSNGMDVFRGDSLTLYKIGGVDFDYFIAGPSPLNVVDQYTLQI 274	
Qy	312 GRPAMPYVAFGHFHCRCWGYRDVNEIETVVDKVAEARIPLVWMTDIDYMDAFKDFLDP 371	
Db	275 GRPAMPYVNSLGFHCRCWGYHNSVVEDVDNYYKAKIPLVDVWDDHGDHDKFTLNP 334	
Qy	372 VHFPLDKNQFVTKLHRNGQRYVFPILDPGINTNKSYGTFIRGMQSNVFIKRNPNVILGSV 431	
Db	335 VAYPRAKLLAFLLKIKHIGMKYIINDPGLGVNASYGTFFQRMADVFIKYGKPFLLAQV 394	
Qy	432 WPGVYVYDFDLDPAAFSFWDSEIKRFDILPIDIGIWMKMEASNFIT-----S 479	
Db	395 WPGVYVYDFDLNPKTVSMWGDSEIKRFDHLDVPLDGLMTDMKNEVSCGLCTIPEGKQCP 454	
Qy	480 APTPG-----STLNDPPYKINNSGGRVPINSKTIPATAMHYGNVTEYNAHNL 527	
Db	455 GEGFGWCCLDCKNITKTRWDPPYKINATGVAPGVFKIATSGHYNGVREYDAHSY 514	
Qy	528 GFLESQATREALVRPATRGPFLLSRSTPAGSKYTAHTWTDGNAARWDDQLYSIPTMLNFG 587	
Db	515 CFSETIATHKLLNVOGKRPFILSRSTPVGSGQVAAHWTDGNDQGTWQSLQVSI STLNFG 574	
Qy	588 LFGEMTGADICGFAESTTEELCCRMQLGAFYFSDHSDARSADTHQELYLWESVAASAR 647	
Db	575 IFGVPWVSDICGFPQPTTEELCNRWIEVGAFYFSDHANYSPROELQWDTVADSAR 634	
Qy	648 TVLGLRYELLPPYTYTMYDANLRGSPARPLSTFPDDVAITYGSSOPLIGRGITWSPVL 707	
Db	635 NALGNRYKILFLTYLNEYAHTCAPARLPFFSFPBYTECYGNSRQFLIGSSFMISPV 694	
Qy	708 QPGSSIVNAYSPRGNVSLSNYTSVVSAGTYVSLSAPPDHIINVHIEGNIIVAMQGEAM 767	

Db	695 EQKTEVEALPPGSGSWTHMDFMTQAVVSKGKRVTLTLPAPLNFVNVHLYQNTILTQOGL 754	
Qy	768 TQAASTPPHLLVVM---SDHVASTGELFDLNG--IEMDIGSPGGKMTLVRFFAESGIN 822	
Db	755 ISKDATTPPSLVIAFPAGASEGVATGKLYLDEDELPKMLG--NGQSTYVDFVASVNG 812	
Qy	823 NLTTSSSEVNRGYAMQWMDKITILGLKRVKIKETTVQKDAGAIKVK-----GL 874	
Db	813 TMRWMSQVKEGKFAUSKGMVIEKVSVLGRAGQGVSEIQINGSPMTKKIEVSSKEHTYVI 872	
Qy	875 GRRTSSNHQGGFFVSVISDLR---QLVGOAFKL 904	
Db	873 GLEDEENK-----SVWVEVGLMLVGVKDFNM 900	
RESULT 8		
Q9S7Y7	PRELIMINARY; PRT; 915 AA.	
ID	Q9S7Y7	
AC	Q9S7Y7	
DT	01-MAY-2000 (TRENBLrel. 13, Created)	
DT	01-MAY-2000 (TRENBLrel. 13, Last sequence update)	
DT	01-MAR-2003 (TRENBLrel. 23, Last annotation update)	
DE	Alpha-xylosidase precursor (ATIG68560/F24J5_10).	
GN	XYLI OR F24J5_20.	
OS	Arabidopsis thaliana (Mouse-ear cress).	
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
OC	Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;	
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.	
OX	NCBI_TaxID=3702;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=cv. Columbia;	
RC	Sampedro J., Sieiro C., Villa T.G., Revilla G., Zarra I.;	
RA	"Cloning and expression pattern of an alpha-xylosidase gene from	
RT	Arabidopsis thaliana."	
RL	Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=cv. Columbia;	
RA	Vyotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,	
RA	Lee J.M., Li J., Gonzalez A., Liu A., Liu K., Vayenberg M., Sakano H.,	
RA	Chin C., Choi E., Chou J., Altati H., Araujo R., Brooks S.,	
RA	Buehler E., Chao Q., Conn L., Conway A.B., Dunn P., Hansen N.,	
RA	Howg B., Huizar L., Khan S., Kim C., Palm C., Rowley D., Shinn P.,	
RA	Walker M., Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;	
RT	"The sequence of BAC F24J5 from Arabidopsis thaliana chromosome 1."	
RL	Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.	
RN	[3]	
RP	SEQUENCE FROM N.A.	
RA	Chen R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banth J.,	
RA	Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayaehizaki Y.,	
RA	Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,	
RA	Kawai J., Lam B., Lee J.M., Lin J., Liu S., Miranda M., Narusaka M.,	
RA	Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,	
RA	Satou C.C., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,	
RA	Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,	
RA	Ecker J.R.;	
RT	"Arabidopsis cDNA clones."	
RL	Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.	
RN	[4]	
RP	SEQUENCE FROM N.A.	
RA	Kim C.J., Chen H., Cheuk R., Shinn P., Bowser L., Carninci P.,	
RA	Chan M.M., Chang C.H., Dale J.M., Hayaehizaki Y., Huan V.W.,	
RA	Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,	
RA	Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,	
RA	Palm C.C., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,	
RA	Tang C.C., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G., Yuan S.,	
RA	Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;	
RT	"Arabidopsis ORF clones."	
RL	Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AF144078; AAD37363.1; -.	
DR	EMBL; ACC08075; AAD49987.1; -.	





```

Db 501 ASGQAPGKYKTIATSAATHYNGVREYDAHSYGFSETHKGLQALEGKRPILTRATF 560
Qy 556 AGSKYTAHTWTGNAARWDDLOYSIPTMLNFGLPFGPMIGADICGF--ASTTEELCCRW 613
Db 561 VSGSHYAAHTGDNKGTWEDLKYSISTVLNFGIFGVPMVGSIDICGFYPAAPPLEELCNRW 620
Qy 614 IQLGAFYPPSRDHSARDTHQELVWESVAASARTVLGLRYELLPPYYTTLWYDANLRGSP 673
Db 621 IQVGAFYPPSRDRHANYSPRQELYQWKSVTYKSRNALCMRYKLLPYLYTLYSYBAHTGAP 680
Qy 674 IARPLSTFFDDVATYGISSQFLGRIMVSPVLQPGSSIVNAYSPRGNWVSISNTSSV 733
Db 681 IVRPLPTFENIPELYELSTQFLVGSVMVSPVLEKAKTKVSALFPFGTWSYLFDMTQVI 740
Qy 734 SVSAGTIVSLSAPDPHINVIHEGNIIVAMQGEAMTTOAARSTFPHLLVM-----SHVAS 789
Db 741 VTKEPHYRSLDAPLHVNVNHLQNTILPMQGGMLTKAEAMTPTTIIIVAPLGASEGVAK 800
Qy 790 TGEFLFONG--IEMDIGPGGKWTLVFAESGINNLTISSEVVNRGYAMSQRVWMDKIT 847
Db 801 -GNLFLDDDELPEMKLG--NGKSTYMDPHATTSGTIVKINSEVQESKYALDKGYIEKVT 857
Qy 848 ILGLKRRVKIKETVYQKDAKAIGK-VKGLGRRTSSH-----NQGGFFVSVISDLRQL--- 897
Db 858 VLINGLIGG--GAFDILVDGSKVEDTSKLEPTEBEHKFDKLEGGHKKSMMLDIKGLELP 915
Qy 898 VQQAFLK 904
Db 916 IGRNFAM 922

RESULT 10
Q9ZP04 PRELIMINARY; PRT; 935 AA.
AC Q9ZP04;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Alpha-D-xyloridase precursor.
OS Tropaeolum majus (Common nasturtium).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Tropaeolaceae; Tropaeolum.
OX NCBI_TaxID=4020;
RN [1]
PP SEQUENCE FROM N.A.
RC STRAIN=cv. tall climbing mixed; TISSUE=Cotyledon;
RA Chombie H.J., Chengappa S., Jarman C., Sidebottom C., Reid G.;
RT "Molecular characterisation of a xyloglucan oligosaccharide-acting
RT alpha-D-xyloridase from the cotyledons of nasturtium (Tropaeolum majus
RT L) seeds";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ31520; CA010382.2; -.
DR InterPro; IPR000322; Glyco_hydro_31.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
KW SIGNAL.
FT SIGNAL. 1 24 POTENTIAL..
FT CHAIN 25 935 ALPHA-D-XYLOSIDASE.
SQ SEQUENCE 935 AA; 104937 MW; 22DB6901E9CE19BD CRC64;

Query Match
Best Local Similarity 40.5%; Score 1969.5; DB 10; Length 935;
Matches 393; Conservative 154; Mismatches 284; Indels 59; Gaps 15;

Qy 18 LPLVLCMVVEGATTSKNDNQ--BAIGYGVQVKNKVDNSTGKSLTALLQIRNSPYGVP 75
Db 16 LPFILLSSSYGNSSHSTPAAPTIGKGYRL--ISIEETPDGQFLGHLQVQKPNKIYGA 73

Qy 76 DIHFLSFSTAFSEEDTLIRITDANNRWEIPNVLPRPPPPPPSLSLQHLKRP-IP- 134
Db 74 DIPLLQVYVGHESQDRLRVHTDAEKORWEYFYNLLPREQPP-----VVEANDREIPG 126

```

```

Qy 135 QNOPTTVLSHPHSDLAFTLPHTTPPGFTIYRKSTHVDLFDATPIPSNPTTFLIYKDOYL 194
Db 127 KNLITVSEIS--GSELIYSYRPRDRFPAVRKSNGETLFNSSSDPSDPFGEMFKDOYL 184
Qy 195 QLSSSLPAQQAHLGLGERTKP-TFOLAHNOQLTLWADIASFNRLDLNLYGSHPPYMDVR 253
Db 195 EISTKLP-KDASLYGLGENTQPHGKLYPNDPYLYTMDVSAINLNADSYSHPPYMDLR 243
Qy 254 SSPMVGSTHGVPLNSGMDVEYTGDRITYKIGGIDLYIPAGRTPEMVLDDQYTKLGR 313
Db 244 NVGGEAYAHAVALLNSGMDVYFRGDSLYTIIGGVDFYFPGAPLDVVDQYTAFIGR 303
Qy 314 PAPMYPAFGHQCRWGYRDVNEIETVVDKYAEARIPLVNMWTDIDYMDAFKDFTLDPVH 373
Db 304 PAPMYFWSLGHQCRWGYHNSLVEDVVERYNKAKIPLDVIWNDDHMDGKKDFLTNKN 363
Qy 374 PPDURKQGFVTKLHRNGORYVPIIDPFGINTKSYGTFFIRGMQSNVFIKRNPNVLSWVP 433
Db 364 YPRPQLLAFLDKSIGIGMKYIVIDPGIANSYGYTQYRGLANDVFIKYEGEPFLAQVMP 423
Qy 434 GPVYVDPFLDPAARSFWDEIKRFRDILPIDGIMDMNEASNFI-----TSAPT-- 482
Db 424 GAVNFEDFLNPKITVDMWGDVEVRHFHELVPDGLMDMNEDELEFLFWMENFPQGCQCTGE 483
Qy 483 -PG-----STLDNPPYKINNSGGRVPINSKTIIPATAMHYGNVTEYNAHNLXGF 529
Db 484 GPGWICCLDCKNITKTRWDEPPYKINATGVQAPIGFKTIATSCYHNGVLEYDAHSIYGF 543
Qy 530 LESQATREALVRPATRPFLLSRSTFAGSKYTAHTWTDGNARWDDLOYSIPTMLNFGIF 589
Db 544 SQTVAHKALQGLEGRPFILTRSTFVSGHYAAHWTGDNQGTWENLRYSISTMLNFGIF 603
Qy 590 GMPMIGADICGFAESTTEELCRMIQLGAFYPPSRDHSARDTHQELVWESVAASARTV 649
Db 604 GVPWGSIDICGFYPOTEELCNRMIEVGAFYPPSRDHANYSPROELQWESVAESARNA 663
Qy 650 LGLRYELLPPYYTLYMYDANLRGSPILRPLSTFFDDVATYGISSQFLGRIMVSPVLPQ 709
Db 664 LGMRYKLLPFLYTLTYEAHIGAPIARLPTFFPNPYECYGVSTQFLGSSLSMISPVLEQ 723
Qy 710 GSSIVNAYSPRGNWVSISNTSSVSVSAGTVSVLSAPPDHINVIHEGNIIVAMQGEAMTT 769
Db 724 GKTEVKALFPFGTWSYLLDMTETVDSKKGQVYTLDAPLHVVNHLVYQNTILPMQGGLLS 783
Qy 770 QAARSTPPELLVW---SDHVASTGELFLONG--IEMDIGPGGKWTLVRFESGINN 823
Db 784 KEARMTPTFLIVTFFPAGATDQAK-GNLFDKDELPEMKLG--NGYSTVVEFYATLNQGA 840
Qy 824 LTISSSEVNRGVAMQRWMDKITILGLKRRVKIKETVYQKDAKAIGKVG 873
Db 841 VKVMSQVQEGKFDALDNGWSIEKVTVLGLS-----NKQVGSLEIDG 880

RESULT 11
Q9LZT7 PRELIMINARY; PRT; 855 AA.
AC Q9LZT7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 95.9 kDa protein.
GN FlcL2 150.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsie.
OX NCBI_TaxID=3702;
RN [1]
PP SEQUENCE FROM N.A.
RA Jordan N., Bangert S., Wiedelmann R., Voss H., Unseld M., Meves H.W.,
RA Rudd S., Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

```



```

Db 397 DAOWNIDYMGCRDFTDFQKFA--SLPSLVDLHKGQHYVILLDPGLSSTSPRGSTW 454
Qy 409 TFRGMSQNFVFK-RNGNPYLGWPGFVYYPDFLPAAARSFWVDEIKRFRDILPIDGIW 467
Db 455 PFDEGLRGLFLNTQQTGLIGQWPGYTAYPDSNTDTHQWLENLQRFHTHPVFDGLW 514
Qy 468 IDNNEASNFITSAP--TPGSTLONPPYK---INNNGRVPINSKTIPTATAMHYGNTEYN 522
Db 515 IDNNEPSNFMGSEBEGCPGELDSPPTPAVLGNS-----LTAKTVCASAEQNASV-HYN 568
Qy 523 AHNLYGFLESQATREALVRPATRGPFLLSRSTFAGSGKYTAHMTGONAAARDLQYSIPT 582
Db 569 LHNLYGLKEAETASALIRIRGKPFVIRSTFPQGRYSGHNLGDNRSQKQMYISPG 628
Qy 583 MLNPLGFMGMIGADICGFAESTTEELCCRWIQLGAFYPSFRDHS-----ARDTHQELY 637
Db 629 MLSPSLFCIPLVGADICGFSGSTSEELCTRWMQLGAFYPSRHHNQNEKAQDPT----- 683
Qy 638 LWESVAASART-----VLGLRYELLPPYYTLMYDANLPGSPIARPLSFTPPDDVATYGIS 693
Db 684 ---AFPSAKTAMKADALLRYSLPLFLYTLFRAHLQGETVARPLFFEFPMVDVATYGLDR 740
Qy 694 QFLIGRIMVSPVLQPGSSIVNAYSPRGNVSLSNYSVSAGTVVLSAPPDHNH 753
Db 741 QFLWQSLVTPVLEFGADSVLGYFPQGVYDF--YTGSSVNSGEMKLSAPLDHLNLH 798
Qy 754 IHENIVAMQEAAMTQQAARSTPHLLVMSDVHSTGBELFLNGIEMDGGPGKRWTLV 813
Db 799 LREGSILPTQKGTISATRGNPLHLVALSTRATANGDLFMDGSLDTFEQGNYSYLV 858
Qy 814 RFAESGI---NNLTISSEVVNGYASQRVWMDKITILGLKRVKIKVYTVQKDAGAIK 870
Db 859 -FNATENIFTSNVLHASTEATD-----VTI-----DAVSFYGVQBPSPSKVL 898
Qy 871 VKGLGRSTS-SHNOGGFFVSVIDRLQLVQAQKLE 905
Db 899 LDQGEKFFSYLDNQ-----VLTVSGLGLVLSQGSFSLQ 930

RESULT 13
Q8BG16 PRELIMINARY; PRT; 953 AA.
AC Q8BG16;
DT 01-MAR-2003 (TremBrel. 23, Created)
DT 01-MAR-2003 (TremBrel. 23, Last sequence update)
DT 01-MAR-2003 (TremBrel. 23, Last annotation update)
DE Glucosidase.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;

Query Match 30.5%; Score 1485; DB 11; Length 953;
Best Local Similarity 38.1%; Pred. No. 4.6e-104;
Matches 331; Conservative 132; Mismatches 299; Indels 106; Gaps 26;

Qy 45 YQVKNKAVDNGTGSKTALLQINSVPYGP-DTHFLSFTASFEEDTLIRFTDANKRR 103
Db 136 YRLLENL---SSTESGYTA--TLTRTSPTFFPKQVLTLLQLEVLMTDSRLHPKIDPASKR 190

```

```

Qy 104 WEIPNEVLPRPPPPPPSLSSQLHLPKPIQONQTTVLSHPHSDLAFTLPHFTTFFGFT 163
Db 191 YEVPLE-----TPRVLQAPSPLYSEFSEPFQVI 221
Qy 164 TYRKSTHVDLPATPIPSNPTFLIYKDYQLQSSLPAAQAHLYLGEGHFKTPTFQLAHN 223
Db 222 VRKLGGRVLANTIVAP-----LFFADQFLQSTSLPSQ--HITGJGHEHLSPLMLSTDW 273
Qy 224 QILTLWADIASFNRLDMLYGSHPFYMDVRSPVMGSHGVFLNSNGMOV-EYTGDRIT 282
Db 274 ARITLWNRDTPP-SQGTNLYGSHFPYLALEDG---GLAHGVFLNSNAMDILOSPALIT 329
Qy 283 YKVGIGIIDLXIEAGRTPEMVLQYTKLIGRAPMPYMAFGHCRMGYRVDNIEIETVD 342
Db 330 WRSTGGIILDVVELGPEPSVVOQLDVGVGFPMPPYMWGLGFHLRCMGYSSTALVRQVE 389
Qy 343 KYAEARIPLEVMWMTDIDYMDAFKDFTLDPVHFHFDKMQOF---VTKLHNRNQRYVPIIDP 399
Db 390 NMTRTHFDLDVQNDLDYMDARROFT-----FNQDSFADFPDMVRELHQDGRYMYVIDP 444
Qy 400 GINT---NKSYGTFIRGMSNVFI-KRNGNPYLGSVMPGPVYYPDFLDPAAARSFWDEIK 455
Db 445 AISSAGPAGSYRVPYDEGLRGVFITNETQPLIGKVPFGTTAFDPFTNPETLDMQDMVS 504
Qy 456 RFRDILPIDGIDWIDNEASNFITSAP--TPGSTLONPPYKINNSGRVPIINSKITIPATAM 513
Db 505 EHAQVDFDGMWLDNNEPSNFRSGSQCCPNNELENPPYVGVVG--ILOAATICASS- 561
Qy 514 HYGNTVEYNAHNLGYFLBSQATREALVRPATRG--PFLLSRSTFAGSGKYTAHMTGONAA 571
Db 562 HQPLSTHYNLHNLGYLTAIASSRALVK--TRGTRPPVISRSTFSGHRVAGHMTGDVRS 619
Qy 572 RWDDLOYSIPTMLNPLGFMGMIGADICGFAESTTEELCCRWIQLGAFYPSFRDHSARDT 631
Db 620 SHEHLAYSVPDITLQNLGLVPLVGADICGFTDSEELCVRWMTQLGAFYPTPMRHNDLNS 679
Qy 632 THQELYLWESVAASA-RTVLGLRYELLPPYYTLMYDANLPGSPIARPLSFTPPDDVATYG 690
Db 680 VQBPYRSETAQQAARWKAFLRYALLPYLTLFRAHVRGDTVARPLFLFEPEDPSTWS 739
Qy 691 ISSQFLIGRIMVSPVLQPGSSIVNAYSPRGNW-----SLSNYTSSV 733
Db 740 VDRQLLWGPALLITPVEFGKTEVTGYFPKGTWYNQMVSVDLSGLTLPSPSSASSFRSAV 799
Qy 734 SVSAGTYVLSAPPDHNHINHEGNIVAMQEAAMTQQAARSTPHLLVMSDVHSTGEL 793
Db 800 Q-SKGQMLTLEAPLDTINVHLREGVILPLOGPSLTFTTESRKQPMALAVALTASGEADGEL 858
Qy 794 FLDNGIEMDGGPGKWTLVRFFAESGINNLTISS--EVMNRGYAMSORVWMDKITILG- 850
Db 859 FWDGDES LAV-LERGAYTLVTFSAK--NNTIVNKLVRVTKEGAEQLR-----EYTVLGV 910
Qy 851 -----LKRVRVKKIKEYTVQKDAGAIKV 871
Db 911 ATAPTQVLSNGIPVSNFTYSPDNKSLAI 938

RESULT 14
Q91245 PRELIMINARY; PRT; 953 AA.
AC Q91245;
DT 01-DEC-2001 (TremBrel. 19, Created)
DT 01-DEC-2001 (TremBrel. 19, Last sequence update)
DT 01-MAR-2003 (TremBrel. 23, Last annotation update)
DE Similar to glucosidase, alpha, acid.
GN GAA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;

```









182 NPTTFLYKDYQLQSSLPAAQAHLYGLGHTKPTFOLAHNOILTLWNADIASFNRLDN 241  
178 NPNTFLFIDQYLLHLSUSLCTAHYHGLGHSKPTFOLAHNOTLTHRAADIPSSNPDVN 237  
242 LYGSHPPYMDVRSSPMVSGTHGVFLNSGMDVEYTGDRITYKVIIGIIDLIFAGRTPE 301  
238 LYGSHPPYMDVRSSPMVSGTHGVFLNSGMDVEYTGDRITYKVIIGIIDLIFAGRTPE 297  
302 MVLDDQYTKLGRAPMPYAFGHOCRWGYRDVNEJETVVKYAEARIPLEVWMTDIDYM 361  
298 QVVEQFTRVGRAPMPYAFGHOCRWGYRDVNEJETVVKYAEARIPLEVWMTDIDYM 357  
362 DAFKDFLDPVHPFLDQMOQFVTKLHRNGORYVPIIDPGINTNKSXGTFRIGMOSNVFIK 421  
358 DAYKDFLDPVHPFLDQMOQFVTKLHRNGORYVPIIDPGINTNKSXGTFRIGMOSNVFIK 417  
422 RGNPNYLGSGVMPGVYVDFDLPAARSFWDEIKRFRDILPDIGIWMIDNEASNFITSAP 481  
418 RNGKPYLGSGVMPGVYVDFDLPAARSFWDEIKRFRDILPDIGIWMIDNEASNFITSAP 477  
482 TPGSTLDNPPYKINNSGGRVPIINSKTIPTAMHYGNVTEYNAHNLGFLSQAATREALVR 541  
478 IPGSTLDNPPYKINNSGGRVPIINSKTIPTAMHYGNVTEYNAHNLGFLSQAATREALVR 537  
542 PATRGPELLSRSTPAGSGKTAHWTGDNAAWDDLOYSIPTMLNFGFLGPMIGADICGF 601  
538 LTKRPPVLSRSTPAGSGKTAHWTGDNAAWDDLOYSIPTMLNFGFLGPMIGADICGF 597  
602 AESTTELCRCRWIQLGAFYFSDHSDARTTHOELYLWESVAASARTVLGLRYELPYYY 661  
598 LGNTTELCRCRWIQLGAFYFSDHSDARTTHOELYLWESVAASARTVLGLRYELPYYY 657  
662 TLMYDANLRGSPARPLSFTFPDDVATYGISOFLGRGMVSPVLOPSSIVNAVSPRG 721  
658 TLMYEAQNLGRIARPLSFTFPDDVATYGISOFLGRGMVSPVLOPSSIVNAVSPRG 717  
722 NWTSLNYSVSSVSAGTYVYSLAPPDHPHNVHIEGNIVAMQGEAMTTOAARSTPPLHLV 781  
718 NWFDFDPTVRSVASTAGRYVTLNAPDHPHNVHIEGNIVAMQGEAMTTOAARSTPPLHLV 777  
782 VMGSHVASTGELFLDNGIEMDIPGSGKWTLPVFPFASGINNLTISSEVNRGYAMSQRW 841  
778 VMSDCGASGELFLDNGIEMDIPGSGKWTLPVFPFASGINNLTISSEVNRGYAMSQRW 837  
842 VMDKITLGLKRRVKIKEYTVQDAGAKVYKLGRTSS--HNQGGFFSVISDLRLQV 899  
838 VIDKVTILGURKGTNGYTVR--TGAVTRKDGSKLKLSTPDRKGEFVVAEISGLNLALG 895  
900 QAFKLEL 906  
896 REFKLVL 902  
RESULT 3  
AGLU HORVU  
ID AGLU HORVU STANDARD; PRT; 877 AA.  
AC Q43763;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase).  
OS Hordeum vulgare (Barley).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
OC Triticeae; Hordeum.  
OX NCBI\_TaxID=4513;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Morex; TISSUE=Aleurone;  
RX MEDLINE=96178863; PubMed=8616248;  
RA Tibbot B.K., Skadsen R.W.;  
RT "Molecular cloning and characterization of a gibberellin-inducible,

RT putative alpha-glucosidase gene from barley.";  
RL Plant Mol. Biol. 30:229-241(1996).  
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-  
CC linked D-glucose residues with release of D-glucose.  
CC -!- TISSUE SPECIFICITY: HIGH LEVELS SEEN IN THE ALEURONE AND SCUTELLUM  
CC AFTER GERMINATION, WHILE LOW LEVELS ARE FOUND IN DEVELOPING SEEDS.  
CC -!- DEVELOPMENTAL STAGE: LEVELS INCREASE STEADILY THROUGHOUT  
CC IMBIBITION REACHING MAXIMUM LEVELS AT DAY 7. DURING GERMINATION,  
CC LEVELS INCREASE FROM DAY 2, REACH MAXIMUM LEVELS AT DAY 3 AND  
CC DECLINE AFTER DAY 5.  
CC -!- INDUCTION: BY GIBBERELLIN A3 (GA).  
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL, U22450; AAB02985.1; -.  
CC PIR, S65057; S65057.  
CC InterPro; IPR00322; Glyco hydro 31.  
CC Pfam; PF01055; Glyco hydro 31; 1.  
CC PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
CC PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.  
CC Hydrolase; Glycosidase; Glycoprotein; Signal.  
CC SIGNAL 1 23 POTENTIAL.  
CC CHAIN 24 877 ALPHA-GLUCOSIDASE.  
CC ACT SITE 437 437 BY SIMILARITY.  
CC FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 338 338 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 391 391 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 570 570 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC SEQUENCE 877 AA; 96933 MW; E3B5C16E4588C492 CRC64;  
Query Match 48.7%; Score 2370; DB 1; Length 877;  
Best Local Similarity 51.2%; Pred. No. 7.5e-151;  
Matches 466; Conservative 147; Mismatches 235; Indels 62; Gaps 15;  
QY 16 VVPLVLVLCVVEGATTSKNDNQGEA---IGYGYQXNA-KVDNST-GKSLTALLQLIRNS 70  
DB 7 LLLCLLCLPAPRLCSSKEGFLAARTVLAVAVTWEGALRAEATGGRSSTG----- 58  
QY 71 PVYGPDIHFLSTASFEEDDTLRIRFTDANNRWEIPEVLPRP-----PPPPSPPP 122  
DB 59 -----DVQRLAVYASLETSRLVRITDADHPREVPODIIPRAPGVDLHDAPASSAP 113  
QY 123 LSSQLHLPKPTPQNPQPTTVLSPHSDLAFTLFTHTPGFTIYKRSHTDVLFDATPIPSN 182  
DB 114 LQG-----RVLSPPAGSDLVLTIV-HASPPRFTVSRSTGDTLFDTP----- 153  
QY 183 PTFELIYKDYQLQSSLPAAQAHLYGLGHTKPTFOLAHNOILTLWNADIASFNRLDN 242  
DB 154 ---GLVFRDKLVTLSALPAGASLYGLGHTKSFRLRHNDSTFLWNADIGASYVDVN 210  
QY 243 YGSHPPYMDVRSSPMVSGTHGVFLNSGMDVEYTGDRITYKVIIGIIDLIFAGRTPE 302  
DB 211 YGSHPPYMDVRAP--GTAHGVLLSSNGMDVLYGGSVTVYKVIIGVLDYFFAGPNPLA 267  
QY 303 VLDQYTKLGRAPMPYAFGHOCRWGYRDVNEJETVVKYAEARIPLEVWMTDIDYM 362  
DB 268 VLDQYTKLGRAPMPYAFGHOCRWGYRDVNEJETVVKYAEARIPLEVWMTDIDYM 327  
QY 363 AFKQFTLDPVHPFLDQMOQFVTKLHRNGORYVPIIDPGINT--NKSXGTFRIGMOSNVF 419  
DB 328 GFKQFTLDPVHPFLDQMOQFVTKLHRNGORYVPIIDPGINT--NKSXGTFRIGMOSNVF 387  
QY 420 IKRNGNPNYLGSGVMPGVYVDFDLPAARSFWDEIKRFRDILPDIGIWMIDNEASNFITS 479



```

Db 388 LKRNGTIFVGNVPCDVFYFDPKHPAAAEFWAREISLFRRTIPVDGLWIDMNEISFYN- 446
Qy 480 APTGSLDNPYPKINNSGRVPIKNSKTIPTAMAHYGNVTYNAHNLVGFLESCATREAL 539
Db 447 -PEPMNALDDPPYRINNDGTRPINNKTVPLAVHYGGVTEYEEHNLFGLEAPARQGV 505
Qy 540 VRPATRGPFLLSRSTFAGSGKYTAHTWGDNAARDLQYSIPTMLNFGFGMPMIGADIC 599
Db 506 LRDTRGRFPVLSRSTFVSGRYTAYWGDNAATWGLRLYSINTMLSGFGMPMIGADIC 565
Qy 600 GFABSTTEELCCRWIQLGAFYFGRDRSHARTTHQELYLWESVAASARTVLGURYELLPY 659
Db 566 GFNGNTTEELCGRWIQLGAFYFGRDRSHARTTHQELYLWESVAASARTVLGURYELLPY 625
Qy 660 YTYLMYDANLRGSPARLPSLTFPDDVATYGISQFLIGRIMVSPVLQPCSSITVNAISP 719
Db 626 FYTYLMEYHMTGAIARLPFFSYPHDVATYGVDRQFLGRGVLSVPLEPCPTVDAYFP 685
Qy 720 RGNWVLSNYSSTVSAGTYVLSAPPDHINVHIHEGNIIVAMQGEAMITTOAARSTPFHL 779
Db 686 AGRWRLYDYSLAVATRTGKHVRLPAPADTVNVHLTGCTILPQQSALITTSRAPRTAFHL 745
Qy 780 LVMSDRHASTGELFLDNGIEMDGGPGKWTLYRFFAESGINN- -LTISSVWNRGYAM 837
Db 746 LVALAEDGTASGYLFLDDGDSPEY-GRSDWSMRYFNKYIPNNKGAIKVKSEVVHNSYAQ 804
Qy 838 SORWMDKITILGKRRVKIKEXTVOKDAGAIKVGGLRRTSSHQGCF-FVSVISDLRQ 896
Db 805 SRTLVISKVLMGHRSPAPKLTWVNSARVEASS-SAGTRYONAGGLGVAVHIGGLSL 863
Qy 897 LVGOAFKLEL 906
Db 864 VVGSEFELKV 873

RESULT 4
AGLU MUCJA
ID AGLU MUCJA STANDARD; PRT; 864 AA.
AC Q92442;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase).
OS Mucor javanicus.
OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
OC Mucor
NCBI_TaxID=51122;
RN [1];
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=IFO 4570;
RX MEDLINE=96271012; PubMed=8830045;
RA Sugimoto M., Suzuki Y.;
RT "Molecular cloning, sequencing, and expression of a cDNA encoding
RL J. Biochem. 119:500-505(1996).
CC -!- FUNCTION: Hydrolyzes not only malto-oligosaccharides but also
CC soluble starch.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-
CC linked D-glucose residues with release of D-glucose
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D67034; BAA11053.1;
DR InterPro; IPR000322; Glyco_hydro_31.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.

```

```

DR PROSITE; PS00707; GLYCOSYL HYDROL_F31_2; FALSE_NEG.
KW Hydrolase; Glycosidase; Glycoprotein_Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 864 ALPHA-GLUCOSIDASE.
FT ACT_SITE 430 430 BY SIMILARITY.
FT CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 364 364 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 406 406 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 466 466 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 500 500 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 734 734 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 864 AA; 98761 MW; CFAB4759DC431403 CRC64;

Query Match 30.0%; Score 1460.5; DB 1; Length 864;
Best Local Similarity 36.4%; Pred. No. 5.4e-90;
Matches 313; Conservative 151; Mismatches 281; Indels 115; Gaps 22;

Qy 9 YICPTLAVVLPVLVWVVEGATTSKNDNGEAGIYGQYGVAKVNDNSTGKSULTAL-LQLI 67
Db 6 FIFVALILITGNVLCQ-----TDATYVSSSAPGY-----KIDGHVKTAEAGLHPLT 53

Qy 68 RNS-----PVYGDPIHLSFTASFEEDTLIRFTDANNRWEIPINEVLPRPPPPS 119
Db 54 LNSRGNKKTGIDTFGXTIKDITVDVEYETEERLHVKSIDKAKKQYLVDPSPGLGFERP--- 110

Qy 120 PPPLSSLOHLPKPIQNQPTTIVLSHPSDLAFTLFTHTPGFTTYRKSTHDVLEDPATPI 179
Db 111 -----QIRHYVSP-----KHSNLDFO-YTAKPFSFKVVRKDDKTTIFDTTNM 151

Qy 180 PSNPTTFLIYKDYQLQLSSLPAAQAHLYGLGEHTKPTFQLAHNQLILTWNADIASFNRD 239
Db 152 P-----LVFEDQYLELSTKVP-EDANIYIGEVTAPE-PRRTHN-VTTLWARD-----NPD 198

Qy 240 ---LNLVYSHPPYMDVRSSPMVGSTHGVFLNSNGMDVEYTGDRITYKVGIGIILYIPA 296
Db 199 DFYRNIYGAHYVYQEVRO---GKAHGALLMNAHGMDVITTEGRITYKVGIGIIFYFA 254

Qy 297 GRT--PEMVLDOYTKLIGRAPMPYMAFGHOCRGYRDVNEIETVVDKYAARIPLEVM 354
Db 255 PKSGKNDLSIAYTDLIGKPMPSHMLGWHHCIRGYPNIDKVEIVTKRYKKEANIPLQTV 314

Qy 355 WTDIDYMDAFKFTLDLPVHFPLDKKQQFVTKLHRNGQRYVPIIDPGINTKNSYGFIRGM 414
Db 315 WVDIDYMEETKDTFTDKVNFQDQRMIGLGEQLKHKQNVVWVDPALISANTTYEYVVRGT 374

Qy 415 QSNVFIR-NGNPYLGSVWPGPVYVDFDLPDPAARSFVDEIKRFRDILPIDGINTDMNEA 473
Db 375 BMDWMIKNADGSDFIGSVWPGFTTTPDMMHPNATKYMNKEIIDFVDMGLGVGLWIDMNEP 434

Qy 474 SNFITSAPTPG-----STLDNPP-----YKI 494
Db 435 ASFCLGSCSGKVDAGNQPYRWYTYEEBOAHHTWEKELKANGNPPGEERNLLVPKVAL 494

Qy 495 NNSGGKVPINSKTIPTAMHYGNVTYNAHNLVGFLESCATREALVRPATR-GPFLLSRS 553
Db 495 NNGAGN-LSEFTVATTALHYGNIPHYDINHLYGHAESHIHQALIKHKNKIRPEVLTRS 552

Qy 554 TAGSGKYTAHTWGDNAARDLQYSIPTMLNFGFGMPMIGADICGPAESTTEELCCRW 613
Db 553 SFGSGSKYGHWTGDNHSPFYLKNSIANILNFMFGVSYSGADYCGFNSDTEELCTRW 612

Qy 614 IOLGAFYPSRDSHARDTTHQELYLWESVAASARTVLGLRYELLPVYTYLMYDANLRGSP 673
Db 613 MEIGAFYPARHNHNNNAKQDPYLVWESTAESRAINTRIYEMLPYFTYTLFRESNRLGUG 672

Qy 674 IARLPSFTPDVATYGISQFLIGRIMVSPVLQPGSSIVNAYSPRGNWVSLNSYTSV 733
Db 673 VWRPLIFEYPAVEELVSDVQLVSGDILLSVLDGKTSVKAQFPFGQYDW--YTHEL 730

Qy 734 SVSAGT-----YVSLSAPPDHINVHIHEGNIIVAMQGEAMITTOAARSTPHLLVMSDHA 788
Db 731 TVDNKSNKKVTVTLDAPLTHIPHIRGGAIIPTKTPKYTVUGETPATPVNLVIALDKKQ 790

```

QY 789 STGELFLDNGIEMDIGPGG 808  
 Db 791 ASGRLYDDGESLEVKSSG 810  
 RESULT 5  
 LYAG HUMAN  
 ID \_LYAG HUMAN STANDARD; PRT; 952 AA.  
 AC P10253; Q14351; Q16302;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Lysosomal alpha-glucosidase precursor (EC 3.2.1.20) (Acid maltase).  
 GN GAA.  
 GE Homo sapiens (Human).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE OF 70-89; 123-145; 204-215;  
 RP 230-249; 332-345; 349-370; 394-409; 480-513; 520-545; 703-719; 726-731  
 RP AND 795-803.  
 RP TISSUE=Placenta, Testis, and Urine;  
 RC MEDLINE=89005058; PubMed=3049072;  
 RA Hoefloot L.H., Hoogveen-Westerveld M., Kroos M.A., van Beeumen J.,  
 RA Reuser A.J.J., Oostra B.A.;  
 RT "Primary structure and processing of lysosomal alpha-glucosidase;  
 RT homology with the intestinal sucrase-isomaltase complex.";  
 RL EMBO J. 7:1697-1704 (1988).  
 RN [2]  
 RP REVISIONS.  
 RA Reuser A.J.J.;  
 RL Submitted (JUN-1990) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90262651; PubMed=2111708;  
 RA Martiniuk F., Mehler M., Tzall S., Meredith G., Hirschhorn R.;  
 RT "Sequence of the cDNA and 5'-flanking region for human acid alpha-  
 RT glucosidase, detection of an intron in the 5' untranslated leader  
 RT sequence, definition of 18-bp polymorphisms, and differences with  
 RT previous cDNA and amino acid sequences.";  
 RL DNA Cell Biol. 9:85-94 (1990).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91037465; PubMed=2268276;  
 RA Hoefloot L.H., Hoogveen-Westerveld M., Reuser A.J.J., Oostra B.A.;  
 RT "Characterization of the human lysosomal alpha-glucosidase gene.";  
 RL Biochem. J. 272:493-497 (1990).  
 RN [5]  
 RP ACTIVE SITE.  
 RX MEDLINE=91310614; PubMed=1856189;  
 RA Hermans M.M.P., Kroos M.A., van Beeumen J., Oostra B.A.,  
 RA Reuser A.J.J.;  
 RT "Human lysosomal alpha-glucosidase. Characterization of the catalytic  
 RT site.";  
 RL J. Biol. Chem. 266:13507-13512 (1991).  
 RN [6]  
 RP CARBOHYDRATE-LINKAGE SITES.  
 RX MEDLINE=93168114; PubMed=8435067;  
 RA Hermans M.M.P., Wisselaar H.A., Kroos M.A., Oostra B.A.,  
 RA Reuser A.J.J.;  
 RT "Human lysosomal alpha-glucosidase: functional characterization of  
 RT the glycosylation sites.";  
 RL Biochem. J. 289:681-686 (1993).  
 RN [7]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=95327152; PubMed=7603530;  
 RA Reuser A.J.J., Kroos M.A., Hermans M.M.P., Bijvoet A.G.A.,  
 RA Verbeet M.P., van Diggelen O.P., Kleijer W.J., van der Ploeg A.T.;  
 RT "Glycosenosis type II (acid maltase deficiency).";  
 RL Muscle Nerve 3:S61-S69 (1995).  
 RN [8]  
 RP VARIANT ASN-91.  
 RX MEDLINE=90365036; PubMed=2203258;  
 RA Martiniuk F., Bodkin M., Tzall S., Hirschhorn R.;  
 RT "Identification of the base-pair substitution responsible for a human  
 RT acid alpha glucosidase allele with lower 'affinity' for glycogen (GAA  
 RT 2) and transient gene expression in deficient cells.";  
 RL Am. J. Hum. Genet. 47:440-445 (1990).  
 RN [9]  
 RP VARIANT GSD-II THR-318.  
 RX MEDLINE=91353580; PubMed=1652892;  
 RA Zhong N., Martiniuk F., Tzall S., Hirschhorn R.;  
 RT "Identification of a missense mutation in one allele of a patient  
 RT with Pompe disease, and use of endonuclease digestion of  
 RT PCR-amplified RNA to demonstrate lack of mRNA expression from the  
 RT second allele.";  
 RL Am. J. Hum. Genet. 49:635-645 (1991).  
 RN [10]  
 RP VARIANT GSD-II LYS-521.  
 RX MEDLINE=91379015; PubMed=1898413;  
 RA Hermans M.M.P., de Graaff E., Kroos M.A., Wisselaar H.A., Oostra B.A.,  
 RA Reuser A.J.J.;  
 RT "Identification of a point mutation in the human lysosomal alpha-  
 RT glucosidase gene causing infantile glycogenosis type II.";  
 RL Biochem. Biophys. Res. Commun. 179:919-926 (1991).  
 RN [11]  
 RP VARIANTS GSD-II ARG-643 AND TRP-725.  
 RX MEDLINE=94004908; PubMed=8401535;  
 RA Hermans M.M.P., Kroos M.A., de Graaff E., Oostra B.A., Reuser A.J.J.;  
 RT "Two mutations affecting the transport and maturation of lysosomal  
 RT alpha-glucosidase in an adult case of glycogen storage disease type  
 RT II.";  
 RL Hum. Mutat. 2:268-273 (1993).  
 RN [12]  
 RP VARIANT GSD-II GLU-645, AND VARIANTS ILE-816 AND ILE-927.  
 RX MEDLINE=93168115; PubMed=8094613;  
 RA Hermans M.M.P., de Graaff E., Kroos M.A., Wisselaar H.A.,  
 RA Willemsen R., Oostra B.A., Reuser A.J.J.;  
 RT "The conservative substitution Asp-645-->Glu in lysosomal alpha-  
 RT glucosidase affects transport and phosphorylation of the enzyme in an  
 RT adult patient with glycogen-storage disease type II.";  
 RL Biochem. J. 289:687-693 (1993).  
 RN [13]  
 RP VARIANT GSD-II GLU-645, AND VARIANTS ILE-816 AND ILE-927.  
 RX MEDLINE=92096118; PubMed=1684505;  
 RA Martiniuk F., Mehler M., Bodkin M., Tzall S., Hirschhorn K., Zhong N.,  
 RA Hirschhorn R.;  
 RT "Identification of a missense mutation in an adult-onset patient with  
 RT glycogenosis type II expressing only one allele.";  
 RL DNA Cell Biol. 10:681-687 (1991).  
 RN [14]  
 RP VARIANTS ILE-816 AND ILE-927.  
 RX MEDLINE=93252406; PubMed=8486380;  
 RA Hermans M.M.P., Svetkey L.P., Oostra B.A., Chen Y.T., Reuser A.J.J.;  
 RT "The loss of a polymorphic glycosylation site caused by Thr-927-->Ile  
 RT is linked to a second polymorphic Val-816-->Ile substitution in  
 RT lysosomal alpha-glucosidase of American blacks.";  
 RL Genomics 16:300-301 (1993).  
 RN [15]  
 RP VARIANT GSD-II VAL-519.  
 RX MEDLINE=95170739; PubMed=7866409;  
 RA Huie M.L., Hirschhorn R., Chen A.S., Martiniuk F., Zhong N.;  
 RT "Mutation at the catalytic site (M519V) in glycogen storage disease  
 RT type II (Pompe disease).";  
 RL Hum. Mutat. 4:291-293 (1994).  
 RN [16]  
 RP VARIANT GSD-II TRP-647.  
 RX MEDLINE=95072571; PubMed=7981676;  
 RA Huie M.L., Chen A.S., Brooks S.S., Grix A., Hirschhorn R.;  
 RT "A de novo 13 nt deletion, a newly identified C647W missense mutation  
 RT and a deletion of exon 18 in infantile onset glycogen storage disease  
 RT type II (GSDII).";  
 RL Hum. Mol. Genet. 3:1081-1087 (1994).  
 RN [17]





```

CC ----- AAC31968.1; --
DR EMBL: AF082188; AAC31968.1; --
DR InterPro: IPR000322; Glyco hydro_31.
DR Pfam: PF01055; Glyco hydro_31.1.
DR PROSITE: PS00129; GLYCOSYL HYDROL F31.1; 1.
DR PROSITE: PS00707; GLYCOSYL HYDROL F31_2; 1.
KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW signal, Cell wall.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 946 GLUCOMYLASE 1.
FT ACT_SITE 462 462 BY SIMILARITY.
FT DOMAIN 519 532 SER/THR-RICH.
FT FT CARBOHYD 51 51 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 97 97 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 244 244 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 373 373 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 393 393 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 406 406 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 437 437 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 505 505 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 570 570 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 772 772 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 801 801 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 895 895 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT CARBOHYD 912 912 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 946 AA; 105804 MW; 5D6B640C4BEF1F70 CRC64;

Query Match 28.5%; Score 1386.5; DB 1; Length 946;
Best Local Similarity 33.7%; Pred. No. 5.5e-85;
Matches 323; Conservative 155; Mismatches 337; Indels 143; Gaps 26;

QY 18 LPLVLCMVVEGATTSKNDN--QGEAIGYGVQVNAKVDNSGKSLTALLQILRNSPVYGP 75
DB 35 VPVELSGVKQLPNHDSAVDANAVAKYSLVNSL---TARGLTGLKLKEATNIYGY 91
QY 76 DIHLSFTASFEEDDTLIRFTDANNRWEIPNEVLRPPPPPPPLSSLSLQHLPKPI-- 133
DB 92 DFEYLNLSVEQSDRLNVHI-----EPTDLTDVFLPEELVV 129
QY 134 -PQNOPTTVLSHPSDLAFLFTHTPPGFTYRKSTHVDLFDATPIPSPTTFLIYKQD 192
DB 130 KPKEGDAKTNFENSOLVFE-YDEEDFGFVLRSSSTREVLFT---KGNP---LVFSN 182
QY 193 YLQSSSLPAQAAH-LYGLGEHTKPTFOAHNQILTLNADIASFNRDLNLYGSHPEYMD 251
DB 183 FIQNTILP--KHSITGLGESIHGSLN-EPGVKTLVANDIAD-PIDGNIYGVHPVYD 238
QY 252 VRSSPMVSTHGVELLSNGMDVEYTGDRITYKVIGGIIDLYIFAGRTPEMVLDOYTKLI 311
DB 239 QRYD--TWTHGVYWRTSIAIQEVVVGTSITWRALSGVIDLYFFSGPPKDVIOQYVEI 296
QY 312 GRPAMPYWGFGHQCRCRGYDNEIETVQYABARIPLEVMKTDIDYMDAFKDFLDP 371
DB 297 GLPAMPQYMALGTHQCRGYDVTESLETWENPKFIDPLETIWSDIDYMDGYKDFNDP 356
QY 372 VHFPLDKMQQFVTKLHRGQRYVPILO-----PGINTKSYGTFIRGMOSNVFIKR-N 423
DB 357 YRFPTDKERKFLDLHNSQHYVPIFDAIYVPPNNATDDOYEPFLHGNESDVFLLKNDP 416
QY 424 GNPYLGSVMPGVVYVDFLDPAPKSFVYDEIKRFRDILPIDGIWIDMNEASNFI----- 477
DB 417 GSLYIGAVWPGYVTPFDPLANNQTEYWNKMKQYERIPFDGIWTDMEVSSFCVSGCGT 476
QY 478 ----- 477
DB 477 DRYEDNPVHPPEVGYSGDYPLGFDKSNASEWKSISEAAAAATKTTTSSASTSIDGK 536
QY 478 -TSAPTPGSLDNPYKYNNSGRVPINSKTIPTAMHYGNVTEYNHNLGFLSQAIR 536
DB 537 NTLAPKGN-INYPYAINNOGDHDLATHAISPNATHADGTVEYDHNHYGLQERAIY 595

```

```

QY 537 EAL--VRPATRPGFLLSRSTFAGSKYTAHTWGDNAARWDDLOYSIPTMLNFGLFQMPMI 594
DB 596 EALLIEHPNKR-PFIIGRSSFAGSKYKYGWGGDNDYADYMYFYIPQALSMLGSGIPFF 654
QY 595 GADIICGFAESTTEELCCRWIQLGAPYFSDRSHARDTHQELYLWESVAASARTVLGLRY 654
DB 655 GVDACGFGNGTDMELCSRWMQLASFPFVYRNHNLGAIPQEPYVWEAVMKATKTSINRY 714
QY 655 ELLPYTYTLMYDANLRGSPAPLSTFTPDVATYGISOFLIGRIMVSPVLPQSGSIV 714
DB 715 SULPYTYTLLHSHVTGIPIMRAFNPQYPSKELAGVDITQFVGDALLVTPVLEPGVNH 774
QY 715 NAYSPRGNWVLSNYT-SSVSAGTYVSLSPAPDHINVHIHEGNIAMQGEAMTTOAAR 773
DB 775 KGIFFGENAVYDFPYTHKKQKFTAGKNETLAAPLGHIPHIKGNIIPTQEPGYTTTTSR 834
QY 774 STPHLLVMSDHVASTGELFLDNGIEMDIGPGGKWTLVRFPAESGINNLTISSEVNR 833
DB 835 KNPPGILLVALDAEGTAGSKLYLDGGSVDV---EALYVDFVASK--NKLVAS---VFG 885
QY 834 GYAMSQRWMDKITILGL---KRYVKIKEYTV--QKDAGAIKVKGLGRRTSSHNOGGF 886
DB 886 EYEVQR--PLANVTILGVDSPEPKVLFNNTVSHKYENGAVYLTLEKPT---KEGAF 938

RESULT 8
MGA_HUMAN
ID MGA_HUMAN STANDARD; PRT; 1856 AA.
AC Q43451;
DT 30-MAY-2000 (Rel. 39, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Maltase-glucoamylase, intestinal [Includes: Maltase (EC 3.2.1.20)
DE (Alpha-glucosidase); Glucoamylase (EC 3.2.1.3) (Glucan 1,4-alpha-
DE glucosidase)].
OS MGAM OR MGA OR MGAML.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Small intestine;
RX MEDLINE=98112863; PubMed=9446624;
RA Nichols B.L., Eldering J.A., Avery S.E., Hahn D., Quaroni A.,
RA Sterchi E.E.;
RT "Human small intestinal maltase-glucoamylase cDNA cloning. Homology
RT to sucrase-isomaltase.";
RL J. Biol. Chem. 273:3076-3081(1998).
RN [2]
RP REVISIONS TO 776; 1049; 1100; 1541; 1612 AND 1811.
RA Nichols B.L., Eldering J.A., Avery S.E., Hahn D., Quaroni A.,
RA Sterchi E.E.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP CHARACTERIZATION.
RC TISSUE=Small intestine mucosa;
RX MEDLINE=89066802; PubMed=3143729;
RA Kaim H.Y., Sterchi E.E., Lentze M.J.;
RT "Structure, biosynthesis, and glycosylation of human small intestinal
RT maltase-glucoamylase.";
RL J. Biol. Chem. 263:19709-19717(1988).
RN [4]
RP SULFATION
RX MEDLINE=88082658; PubMed=3121301;
RA Danielson E.M.;
RT "Tyrosine sulfation, a post-translational modification of microvillar
RT enzymes in the small intestinal enterocyte.";
RL EMBO J. 6:2891-2896(1987).
CC -!- FUNCTION: MAY SERVE AS AN ALTERNATE PATHWAY FOR STARCH DIGESTION
CC WHEN LUMINAL ALPHA-AMYLASE ACTIVITY IS REDUCED BECAUSE OF
CC IMMATURITY OR MALNUTRITION. MAY PLAY A UNIQUE ROLE IN THE
CC DIGESTION OF MALTED DIETARY OLIGOSACCHARIDES USED IN FOOD

```



Db 915 KNGVPSQTSPTVTYDNLKVAIITDILLGDAYTVE 952

## RESULT 9

AGLU SCHPO STANDARD; PRT; 969 AA.  
AC Q9C0Y4;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase).  
GN AGL OR SPAPB24D3.10C.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A., SEQUENCE OF 25-36; 175-194; 375-395 AND  
RP 427-451 AND MUTAGENESIS OF ASP-481; GLU-484 AND ASP-647.  
RX MEDLINE=21195240; PubMed=11298744;  
RA Okuyama M., Okuno A., Shimizu N., Mori H., Kimura A., Chiba S.;  
RT "Carboxyl group of residue Asp647 as possible proton donor in  
RT catalytic reaction of alpha-glucosidase from Schizosaccharomyces  
RT pombe.";  
RL Eur. J. Biochem. 268:2270-2280(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RP STRAIN=972;  
RX MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Scouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren R., Whitehead S.,  
RA Woodward J., Volkhardt G., Aert R., Robben J., Grynoprez B.,  
RA Weltjens I., Vansireels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Fumelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzor A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
RT "The genome sequence of Schizosaccharomyces pombe.";  
RL Nature 415:871-880(2002).

CC -i- FUNCTION: Hydrolyzes malto-oligosaccharides, but has a low

CC activity toward soluble starch.

CC -i- CATABOLIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-

CC linked D-glucose residues with release of D-glucose.

CC -i- SUBCELLULAR LOCATION: Secreted.

CC -i- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; AL590582; CAC36906.1; -

CC DR EMBL; AB045751; BAB43946.1; -

DR GeneDB SPombe; SPAPB24D3.10C; -  
DR InterPro; IPR000322; Glyco\_hydro\_31.  
DR Pfam; PF01055; Glyco\_hydro\_31; 1.  
DR PROSITE; PS00129; GLYCOSYL HYDROL\_F31\_1; 1.  
DR PROSITE; PS00707; GLYCOSYL HYDROL\_F31\_2; FALSE\_NEG.  
KW Hydrolase; Glycosidase; Glycoprotein; Signal.  
FT SIGNAL 1 24  
FT CHAIN 25 969 ALPHA-GLUCOSIDASE.  
FT ACT\_SITE 481 481 NUCLEOPHILE.  
FT ACT\_SITE 484 484  
FT ACT\_SITE 647 647  
FT CARBOHYD 37 37 PROTON DONOR.  
FT CARBOHYD 67 67 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 99 99 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 139 139 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 146 146 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 245 245 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 249 249 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 331 331 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 406 406 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 429 429 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 462 462 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 470 470 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 520 520 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 523 523 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 529 529 O-LINKED (POTENTIAL).  
FT CARBOHYD 543 543 O-LINKED (POTENTIAL).  
FT CARBOHYD 589 589 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 648 648 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 801 801 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 810 810 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 821 821 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 885 885 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 915 915 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 934 934 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 942 942 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 954 954 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 966 966 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CONFLICT 30 30 L -> F (IN REF. 2).  
FT CONFLICT 220 220 P -> A (IN REF. 1).  
FT CONFLICT 507 507 T -> V (IN REF. 1).  
FT CONFLICT 566 566 D -> N (IN REF. 1).  
SQ SEQUENCE 969 AA; 108686 MW; F3122E2CFA551C25 CRC64;  
Query Match 28.0%; Score 1360.5; DB 1; Length 969;  
Best Local Similarity 33.2%; Pred. No. 3.1e-83;  
Matches 316; Conservative 154; Mismatches 320; Indels 163; Gaps 26;  
QY 44 GYQVKNAKVDNSTGKSLTALLQIIRNSP--VYGPDIHFLSFTASFEEDTLIRIRFTDANN 101  
DB 62 GYQAVNI---SESQNGVTAYLALL-GEPCYAYGYDYLFLNVTYEADRVHISIKDANN 117  
QY 102 RWEIFNEVLPRPPPPSPPLSSLOHLKPKIPQOQPTTTVLSHPHSDLAFTL-FHTTFF 160  
DB 118 TQFQFTSKOLWDAPLYSP-----SYNTNLLNYFSYNANPF 154  
QY 161 GFTYRKSTHVDLFDATPIPSNPTFLYKDQYLQLSLPAQOAHLYGLGHTKPTFQL 220  
DB 155 EFWVTRKSDGEVLFDTR-----QKLVFEDQYIELTNNM-VENYNLYGLAE-TIHGRL 206  
QY 221 AHNQILTNWADIASFNRDNLVYGHFYMVDR-----SSPMVGSTHGVFLNS 269  
DB 207 GNNLTRTFWANDEPS-PVDQNMVGHSPHYLFQRYKADGINSTLNETTTSSSHGVMLTA 265  
QY 270 NGMDVEYTGDRITVKYIGGIIDLYIFAGRT--PEMWLDQYTKLGRPAMPYPAWFGFHC 327  
DB 266 NGMDVLLRQDYQLQVRMLGGVLDLFFVYSGSTSPKETVKQFVQSIGKPAHQYWTGLGHSC 325  
QY 328 RWGYRDVNEIYVDKYABARIPLEVNMWTDIDYMDAFKDFLDPVHPFLDKMQQOQVTKLH 387  
DB 326 RWGYTNTITEIMDVQRNYIDADIPVETFWSDIDYMEKYRDFTDVDPVYSKSDMQTFFSDLV 385

```
QY 388 RRGQRYVPIIDPGI-----NTKSYGTFIRGQSNVFIKR-NGNPYLGVSVPQGVYPV 439
Cc
Cc
Cc
Db 386 SNHQHVYPIIDAIAAYANPNYHDDSYYPYAGVEXDIPLKPNGSIIYIGAVWPGETAFP 445
Cc
Cc
Cc
QY 440 DFLDPAARSEFWDEIKRF-----RDLPIDGHWIDMNEASFNITS-----479
Cc
Cc
Cc
Db 446 DFTNPQVDVYWKDCLINLTVAFGSNGTVPPSGIWTDMNEPSSFCVSGSGSAMIDLNPASP 505
Cc
Cc
Cc
QY 480 -----APT-----PGSTL 487
Cc
Cc
Cc
Db 506 LTGISKOYSIPEGFNVSNTVEYSASVSLSNYATATSSVFOIVSPTATPLGKRPDNYNI 565
Cc
Cc
Cc
QY 488 DNPYPKINSGRVPINSKTIPTAMHYGNVTVEYNHNLGYFLESQATREAL--VAPATP 545
Cc
Cc
Cc
Db 566 DWPPYALNNEQGNHDIANHIVSPNATTHDGTQRYDIPNMYGYGETKVSAAULTQISPNER 625
Cc
Cc
Cc
QY 546 GPELLSRSTFAGSGKTAHWGTGNAARWDDLOQVSIPTMLNFGLFPMPIGADICGPAEST 605
Cc
Cc
Cc
Db 626 -PFILSRSTFLGSGVYGAHWLGDHSLWSNMFISGIMVFNMMGIPMVGADVCGFLGDS 684
Cc
Cc
Cc
QY 606 TEELCCRWIOLGAFYPSRDHSARDTHOELYLWESVAASARTVLGLRYELLPYYVTLMY 665
Cc
Cc
Cc
Db 685 DEELCRWAWGAFSPYRNHNNIYQISOEPTWSSVAASRAMYIRYSLPYWTITMA 744
Cc
Cc
Cc
QY 666 DANLRGSIAPRLSFTFPDDVATYGISSQFLIGRMVSPVLQFGSSIVNAYSPRGN---722
Cc
Cc
Cc
Db 745 KASQDCTPALRALFVEFPNDPLADVDQRPMVGDLSLLVTPLEPNVEYVQGVFGDNSTV 804
Cc
Cc
Cc
QY 723 WVSLSNYTSVSAGTYVLSAPPDHNHVIHEGNIVAMQGEAMTQARSTPFHLVV 782
Cc
Cc
Cc
Db 805 WYDWNHNTIETR-QYENVTLYAPLEHINVAIRGGSVLPVMOQPSLTYESRQNPENLLVA 963
Cc
Cc
Cc
QY 783 MSDHVASTGELFLDNGLEMDIGFGGKWTLL-VREFFAESGINNLTISSEVNVNRYAMSQRW 941
Cc
Cc
Cc
Db 864 LDRGSGATGELYDDGVSIELNA-----TLVSFTTSDG-----VLSAVPTGSYEVSQ--911
Cc
Cc
Cc
QY 842 VMDKITILKRR-----VKIKEYTVOKDAGAKVKGLGRTSSHQGGF 886
Cc
Cc
Cc
Db 912 PLANVILGLTESPSSITLNGQNVSSFOYSNDTEELLITGLQNITSS---GAF 961
Cc
Cc
Cc
RESULT 10
SUIS_RABIT
ID_SUIS_RABIT STANDARD; PRT; 1826 AA.
AC P07768;
DT 01-AUG-1988 (Rel. 08, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sucrase-isomaltase, intestinal [Contains: Sucrase (EC 3.2.1.48);
DE Isomaltase (EC 3.2.1.10)].
GN SI.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=86245068; PubMed=3755079;
RA Hunziker W., Spiess M., Semenza G., Lodish H.F.;
RT "The sucrase-isomaltase complex: primary structure, membrane-
RT orientation, and evolution of a stalked, intrinsic brush border
RT protein."
RL Cell 46:227-234 (1986).
RN [2]
RP PRELIMINARY SEQUENCE OF 1-37 AND 1007-1014.
RX MEDLINE=83105704; PubMed=7152027;
RA Sjoestrom H., Noren O., Christiansen L.A., Wacker H., Spiess M.,
RA Bigler-Meier B., Rickli E.E., Semenza G.;
RT "N-terminal sequences of pig intestinal sucrase-isomaltase and pro-
RT sucrase--isomaltase. Implications for the biosynthesis and membrane
RT insertion of pro-sucrase--isomaltase."
RL FEBS Lett. 148:321-325 (1982).
```



FT CARBOHYD 1402 1402 N-LINKED (GLCNAC. . .) (POTENTIAL) .  
FT CARBOHYD 1534 1534 N-LINKED (GLCNAC. . .) (POTENTIAL) .  
FT CARBOHYD 1571 1571 N-LINKED (GLCNAC. . .) (POTENTIAL) .  
FT CARBOHYD 1747 1747 N-LINKED (GLCNAC. . .) (POTENTIAL) .  
FT CARBOHYD 1762 1762 N-LINKED (GLCNAC. . .) (POTENTIAL) .  
FT CARBOHYD 1798 1798 N-LINKED (GLCNAC. . .) (POTENTIAL) .  
SQ SEQUENCE 1826 AA; 210008 MW; 6840D03955A45BE5 CRC64;  
  
Query Match 27.98; Score 1355.5; DB 1; Length 1826;  
Best Local Similarity 35.14; Pred. No. 1.6e-82;  
Matches 320; Conservative 157; Mismatches 313; Indels 121; Gaps 28;  
  
Qy 43 YGVQVKNKVDNSTGKSLTALLQLIRNSPVYGDPIHFLSPTAFSEDDTLRIKFTDANNR 102  
Db 111 HGVNVEGM---TTSTGLEARLNKSTPTLFGDINNVLTTESQTANRLRFLKLTDPNKK 167  
Qy 103 RWEIPNEVLPRPPPPSPPLSSLOHLPKPIPNQNPPTTVLSHPHSDLAFTLPHFTTFFGF 162  
Db 168 RYEVPHQVTEFAGPAATETLYDVQ-----VTEN-----PFSI 200  
Qy 163 TIYRKSTHDLVDPATPSPNPTFLIYKQOYLQSSSLPAQAHLXCLGEHTKPTQLAH 222  
Db 201 KVIKSNRNLIFDSSIGP-----LVSDQYLQISTLSPSE--YMYGFGEHVHKKRFR--H 250  
Qy 223 NQILTLWNADIASFNRDL-----NLYGSHPFVMDVRSSPMVSGTHGVFLNLSNGMDVE 275  
Db 251 DLYWKTW---PIFTRQHTDDNNNLYGHQTFMCIEDT--TGKSGFVFLMNSNAMEIF 304  
Qy 276 YTGDRIT-YKVGIGIIDLIPAGTBPMLVDQYTKLIGRPAPMPYNAFGPHQCRWGRDV 334  
Db 305 IQPTPIVYRVIGILDYFLGDTPEQVVQYQELIGRPAMPAYMSLQPSLRMYNSL 364  
Qy 335 NEIETVVDKYAEARIPLVNMVTDIDYDAEKDFTLDPVHFLPKMOQVFTKLHNSGRVY 394  
Db 365 DVKVEVRRREALIPDPTQVSDIDYEMDKDFTYDRVAY--NGLPDPVQDLHDHQQKYV 422  
Qy 395 PILDPGINTNK-----SYGTFIRMQSNVFIKRG--NPYLGSVWPFPVYYPDFLPPAAR 447  
Db 423 IILDPAISINRRAGEAYESYDRGNAQVWVNESDGTTPVGEWPGDVTYVPDFTSPNCI 482  
Qy 448 SFVWDEIKRFRDILPIDGIWIDNEANFI--TSAPTPGSTLONPPYKINNNGSRVP--- 502  
Db 483 EWMANEENI FHOEYVNOGLIDMNEVSFVQSGNKGNDNLNPPY--PDIV 534  
Qy 503 ---INSKTIPATAMHYGNVTEYNAHNLVGFLESQATREALVR--PATRGPFLLSRSTFAG 557  
Db 535 DKLMYSTKLCHMSVQYWG-KQYDVHSLYGYSMATATERAVERVPNKRK-FILTRSTFAG 592  
Qy 558 SGKYTAHTWGDNAAARDLQVSIPTMLNFGLFPMITGADICGFAESTTEELCCRWIQLG 617  
Db 593 SGRHAHMLGDNATATWEQMEWSITGMLEFGLFGPLVGVADICGLFAETTEELCCRWIQLG 652  
Qy 618 AFYFSDHSHARDTHOELYLW-----ESVASARTVLGLRYELLPPYYTLMYDANLGRSPI 674  
Db 653 AFYFSDHSHARDTHOELYLW-----ESVASARTVLGLRYELLPPYYTLMYDANLGRSPI 674  
Qy 675 ARPLSFTFPDDVATYGISOFFLIGRGVWSPVLQPGSSIVNAYSPRGNWVSLNSNYTSSVS 734  
Db 713 ARPVLHFEYEDTNSWEDREFLWGPALLIITPVLTOGAETVSATIPDAVWYDYE--TGAKR 770  
Qy 735 VSAGTYVLSLAPDHINVHIHEGNIVAMQGEAMTTOAARSTPHLLVMSDHSVASTGELF 794  
Db 771 PWRQORVEMSLPADKIGLHRLGGVYIPIQPAVTTTASRMNPLGLIALNDNDTAVGQFF 830  
Qy 795 LDNGIEMDGGPGGKWTLVRFPAESGINNLTISSEVNRGVAMSQRVWMDKITILGLKRR 854  
Db 831 WDDGETKDT-VQNDNYILYTFVAVSNNLNITCTHLYSEGTTLA----FQTIKILGVET 885  
Qy 855 VKIKEYTVQKDAGAIKVGGLGRRTSSH-----NQGGFFVSVISDLRVLQVQAFKLE- 905  
Db 886 V--TQVTVAEEN-----NQSMSTHSNFTYDPSNQ-----VLLIENLNFNLGRNFRVQW 930  
Qy 906 ----LEFEGAT 912

Db 931 DOTFLESEKIT 941  
  
RESULT 11  
AGLU ASPNG  
ID AGLU ASPNG STANDARD; PRT; 985 AA.  
AC P56526; O13451;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase).  
GN AGLA.  
OS Aspergillus niger.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
OX NCBI\_TaxID=5061;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GN-3;  
RX MEDLINE=97308536; PubMed=9165762;  
RA Nakamura A., Nishimura I., Yokoyama A., Lee D.-G., Hidaka M.,  
RA Masaki H., Kimura A., Chiba S., Uozumi T.;  
RT "Cloning and sequencing of an alpha-glucosidase gene from Aspergillus  
RT niger and its expression in A. nidulans.";  
RL J. Biotechnol. 53:75-84(1997).  
RN [2]  
RP SEQUENCE OF 26-252 AND 267-985.  
RC STRAIN=GN-8;  
RX MEDLINE=93005089; PubMed=1368849;  
RA Kimura A., Takata M., Sakai O., Matsui H., Takai N., Takayanagi T.,  
RA Nishimura I., Uozumi T., Chiba S.;  
RT "Complete amino acid sequence of crystalline alpha-glucosidase from  
RT Aspergillus niger.";  
RL Biosci. Biotechnol. Biochem. 56:1368-1370(1992).  
CC -!- FUNCTION: Hydrolyzes malto-oligosaccharides, but has a low  
CC activity toward soluble starch.  
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-  
CC linked D-glucose residues with release of D-glucose.  
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D45356; BAA23616.1; -;  
CC GlycoSuiteDB; P56526; -;  
DR InterPro: IPR00322; Glyco\_hydro\_31.  
DR Pfam: PF01055; Glyco\_hydro\_31; 1.  
DR PROSITE: PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
DR PROSITE: PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.  
KW Hydrolase; Glycosidase; Glycoprotein; Signal.  
FT SIGNAL 1 25  
FT CHAIN 26 985 ALPHA-GLUCOSIDASE.  
FT ACT\_SITE 490 490  
FT CARBOHYD 36 36 O-LINKED (POTENTIAL) .  
FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) .  
FT CARBOHYD 143 143 N-LINKED (GLCNAC. . .) .  
FT CARBOHYD 218 218 N-LINKED (GLCNAC. . .) .  
FT CARBOHYD 347 347 N-LINKED (GLCNAC. . .) .  
FT CARBOHYD 422 422 N-LINKED (GLCNAC. . .) .  
FT CARBOHYD 506 506 N-LINKED (GLCNAC. . .) .  
FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) .  
FT CARBOHYD 537 537 N-LINKED (GLCNAC. . .) .  
FT CARBOHYD 545 545 O-LINKED (POTENTIAL) .  
FT CARBOHYD 550 550 O-LINKED (POTENTIAL) .  
FT CARBOHYD 559 559 O-LINKED (POTENTIAL) .  
FT CARBOHYD 560 560 O-LINKED (POTENTIAL) .  
FT CARBOHYD 561 561 O-LINKED (POTENTIAL) .

```
FT CARBOHYD 562 O-LINKED (POTENTIAL) .
FT CARBOHYD 601 N-LINKED (GLCNAC. . .)
FT CARBOHYD 623 N-LINKED (GLCNAC. . .)
FT CARBOHYD 835 N-LINKED (GLCNAC. . .)
FT CARBOHYD 881 N-LINKED (GLCNAC. . .)
FT CARBOHYD 899 N-LINKED (GLCNAC. . .)
FT CARBOHYD 957 N-LINKED (GLCNAC. . .)
FT CARBOHYD 970 N-LINKED (GLCNAC. . .)
FT VARIANT 27 TT -> LL (IN STRAIN GN-8) .
FT VARIANT 42 N -> A (IN STRAIN GN-8) .
FT VARIANT 929 N -> M (IN STRAIN GN-8) .
SQ SEQUENCE 985 AA; 108913 MW; 9A18772AEB2E0927 CRC64;

Query Match 27.6%; Score 1339.5; DB 1; Length 985;
Best Local Similarity 32.6%; Pred. No. 8.1e-82;
Matches 324; Conservative 156; Mismatches 306; Indels 209; Gaps 29;

QY 7 PRYICPTLAIVLPLVLCMVVEGATTSKN-----DNQGEAIGYGYQVKNAKVDNSTOKS 59
Db 33 POFTIPASADV-----GAQLIANIDDPQADAQSVCPGY----KASKVQHSN-RG 77
QY 60 LTALLOLI-RNSPVGPDHIFLSFTASPEEDTLIRF-----TDANRRWEIPVLP 114
Db 78 FTASQLAGRPNVGTDESITLSVEYQDSRLNIQILPHTHVDSTNASWFLSEN--- 134
QY 115 PPPSPPLSSQLPKPIQPOPTTVLSHPSDLAFTLHTTTPFGFTIYRKSTHDLVLF 174
Db 135 ----VPRPKASLN-----ASVSQSDLFVMSNEPSPFNKVIKATGDLAF 175
QY 175 DATPIPSNTFTLIYKQVLQSSLPAAQAAHLGLGEHTKPTTFOALNQILTLNADIA 234
Db 176 -----STECTVLVYENQFIEFTALP-BEYNLYGLGEHI-TQFRLQRNANLTIYPSDDG 227
QY 235 SFNRDLNLYGSPFYVDYR-----SSPMVGSTHGVFLNLSNMGD 273
Db 228 T-PIDONLYGQHPFLYLDTRYKGRQNGSYIPVKSEADASQDYISLSHGVLNRSHGLE 286
QY 274 VEYTGDRITYKIGGIDILYIFAGRTPEVMDQY-TKLIGRPAPYWFAGPHQCRWYR 332
Db 287 ILLRSOKLWRLTGGIDILTFYSGPAPADVTRYLTSTVGLPAMQYNTLGHQCRWYN 346
QY 333 DYNEIETVVDKYAEARIPLEVMTDIDYMDAPKFTLPDPVHPFLDMQOQVTKLHRNQOR 392
Db 347 NWSDLADVANFEKEIPELEYITWDIDYMHGVRNFENDQHRFSYEGDEFSLKHSGRY 406
QY 393 YYPILDPGI-----NTNKSYGTFIRGMSQNVFIKR-NGNPVLGSLVWPCVYYPFLDPA 445
Db 407 YVPIVDAALYIPNENASDAYATYRGAADDVFLKPNPDGSLYIGAVMPGYTVFPDWHHPK 466
QY 446 ARSFVDEIKRFRDILPIDGIWDMNEASNF----- 476
Db 467 AVDFWANELVINSKVAFGVGYDMSEVSFCVSGCGTGNLTNPAHPSFLLPGEPIII 526
QY 477 -----ITSA-----PTPG-STLDNPPYKNN 496
Db 527 YDYPEAFNITNATEASASAGASSQAATATTTSTSVSLRTTTPGVRNVEHPPIVNH 586
QY 497 SGRVPINSKTIPTAMHYGNVTEYNAHNLGFLESQATREAL--VRPATRGVPELLSRST 554
Db 587 DEGHDLVHVASPNATHVGVDEEDVHGLYGHQGLNATYQGLLEVHWSHRRPFIIGRST 646
QY 555 FAGSGKYTAHTGDNAARDDILQYSIPTMLNFGLCFQPMIGADICGFAESTTEELCCRWI 614
Db 647 FAGSGKMGWGGDNYSKWMSMYSTISQALSLSLFGIPMGADTCFGNGSDEELCNRMW 706
QY 615 QLGAFTFGRDHSARDTHQELVYVESVAASARTVLGLRYELLPPYYTLMYDANLRGSP 674
Db 707 QLSAFFFFVRNHNELSTIPEQPYRASVIEATKSMRIYAILPYPTLFDLAHTTGSTV 766
QY 675 ARPLSTFFDDVATYGISQFLIGRIMVSPVLOPSSIVNAYSP---RG----NWVSLS 727
Db 767 MRALSWEFFNDPTLAAVETQFMVGPVIMVVPVLEPLVNTVKGVFGVGHGVYDMW---- 822
```

```
QY 728 NYT-SSVSAGTVLSLSAPPDHNVHIEGNIVAMQGEAMTTOARSTPFHLLVMSDH 786
Db 823 -YQAAVDAKPGVNTTISAPLGHIPYVRGNGILPMQEPALITREARQTWALLAALGSN 881
QY 787 VASTGBELFDNGIEMDIMGPGGKWTLLVRFPAESGNNLTISSE-VNVRGYAMSO---RW- 841
Db 882 GTASGQLYLDG-----ESIYPNATLHVDFTASSRSLRSSAQRW 922
QY 842 ---VMDKITILGLKRVKIKKEYTVQKDAGAIKVG 873
Db 923 ERNPLANVTVLG-----VNKEPSAVTLNG 946

RESULT 12
AMYG DEBOC
ID AMYG DEBOC STANDARD; PRT; 958 AA.
AC P22861; Q92336;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Glucoamylase 1 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)
DE (1,4-alpha-D-glucan glucohydrolase).
GN GAMI.
OS Debaryomyces occidentalis (Yeast) (Schwanniomyces occidentalis).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Debaryomyces.
OX NCBI_TaxID=27300;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=ATCC 26076;
RC MEDLINE=91071592; PubMed=1979298;
RA Dohmen R.J., Strasser A.W.M., Dahleens U.M., Hollenberg C.P.;
RA "Cloning of the Schwanniomyces occidentalis glucoamylase gene (GAM1)
RT and its expression in Saccharomyces cerevisiae.";
RL Gene 95:111-121 (1990).
RN [2]
RP SIMILARITY TO OTHER FAMILY 31 MEMBERS.
RX MEDLINE=92077121; PubMed=1743281;
RX Naim H.Y., Niermann T., Kleinbans U., Hollenberg C.P.,
RX Strasser A.W.M.;
RT "Striking structural and functional similarities suggest that
RT intestinal sucrase-isomaltase, human lysosomal alpha-glucosidase and
RT Schwanniomyces occidentalis glucoamylase are derived from a common
RT ancestral gene.";
RL FEBS Lett. 294:109-112 (1991).
CC -!- FUNCTION: THIS GLUCOAMYLASE HAS A SPECIFICITY TOWARD BOTH
CC ALPHA-1,4 AND ALPHA-1,6 LINKAGES.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; M60207; AAA33923.1; -.
CC PIR; JN0102; JN0102.
CC InterPro; IPR000322; Glyco_hydro_31.
CC Pfam; PF01055; Glyco_hydro_31.1.
CC PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
CC PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
CC Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
CC Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 958 GLUCOAMYLASE 1.
FT ACT_SITE 470 470 BY SIMILARITY.
FT DOMAIN 26 41 SER-RICH.
FT DOMAIN 530 542 SER/THR-RICH.
```

FT	CARBOHYD	61	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	78	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	107	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	197	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	403	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	416	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	513	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	580	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	602	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	813	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
FT	CARBOHYD	907	N-LINKED (GLCNAC. . .)	(POTENTIAL)	
SQ	SEQUENCE	958 AA; 106507 MW; 47938DB9BC308260 CRC64;			
Query Match					
Best Local Similarity 33.28; Pred. No. 7.9e-81; Length 958;					
Matches 318; Conservative 152; Mismatches 315; Indels 173; Gaps 32;					
QY	34	NDN--QGEAIGYGVQVKNVAKVDNSTGKSLTALLQLIRNSPVYDPDHFSLFTASFEEDDT	92		
Db	61	NDSAVDANAAGYDLNVV---TNTPRGLTGILKKEATNIYGVDFDNLNLTVEYQADTR	117		
QY	92	LRIRFTDANRRWEIPNVLPRPPPPPLSLQHLPL-----KPIPNQPTTVLSPH	146		
Db	118	LNWHI-----EPTDLSDFVLPEHLVVKPLVEGDAQS--YNFD	153		
QY	147	HSDLAFTLFTTPPGFTYRKSTHDLVDFATPIPSNPTFLIYKDOVLQLSSSLPAQQA	206		
Db	154	NSDLVFE-YSNTDSEFVIRSSKEVLFST---KGNP---LVFSNQFIQNSSLPKNHV-	205		
QY	207	LYGLGEH-----TKPTQLAHNQLITLWNADIASFNRLNLYGSHPPYMDVRSSPMVGST	261		
Db	206	ITGLGESIHGLVNEP-----GSVKTLFANDVGD-PIDGNIYGVHPVYLDORVD--TEIT	256		
QY	262	HGVFLNSNGMDVEYTGDRITYKVIIGIIDLIFAGTPEWLDQYKTLGRAPMPYWA	321		
Db	257	HAVTWRTSAIQEVLIGESITWALSVIDLYFSGPTPKDAIQYVKEIGLFAFOPYWS	316		
QY	322	FGHCQRGWYDVNEIETVYDKAEARIPLEVWMTDIDYMDAFKDFTLDPVHPFLDKMQQ	381		
Db	317	LGYHOCRWGYDTIEKSEVENFKFNIPILETISWDIDYDSYKDFYDHPHFDEYRK	376		
QY	382	FVTKLHRNGRYVPILD-----PGINTKNSGYTFIRGQSNVFIKR-NGNPYLGSWP	433		
Db	377	FLOBELHKNNGHYFILDAAIYVFNPNATNEYPFHGNETDVLKNDPGSLYIGAVWQ	436		
QY	434	GPVYYPDFLPAARSFWDEIKRPDLIPIDGIMIDMNEASNF-----	476		
Db	437	-VTLFSRFLS-RKHSMDOKVIKWYELTPFDGIWADNMNEVSSFCVSGCTGKYFENPAYP	494		
QY	477	-----ITSAPTGSTLD-----	488		
Db	495	PFTVGSKATSYPVGFDVSNASEMKSIOSSISATAKTSTSSVSSSSTIDYMNLTAPGKG	554		
QY	489	---NPPYKINSGRVPINSKTIPTATAMHGVNTEYNAMNLYGFLESOATREAJ--VRPA	543		
Db	555	KINYPYAIYVMQDSDLAHVASPNATHADGVTEYDIHNLGYLOBNATYHALLEVEFN	64		
QY	544	TRGPFLLSRSTFAGSGKYTAHTWTDGNARWDDLOYISPTMLNGLFGMPMIGADICGEAE	603		
Db	615	KR-PFMIISRSTFPAGKWTGWGDDNTADWAYFSIPQAFSGIAGLPFFGADVCGFNG	673		
QY	604	STTEELCCRIQLGAFYPPFRSDHARDTHQELYLWESVAASARTVLGLRYELLYPYTL	663		
Db	674	NSDSELCSRMQLGSGFFPFRNHNLYGAIDQEPYVWESVAEATRTSMALYLLPYTYTL	733		
QY	664	MYDANLKGSPARPLSTFTPDVDATYGISQFOLIGRIMVSPVLQPGSSITWVAYSERG--	721		
Db	734	LHESHTTGLPIRAFSAQGFNDRSLSGVDNQFFVGVDGLVTVTVLEPGVDKVGKVFPGAGK	793		
QY	722	-----NWVSLSNYTS-SVSVSAGTYVLSAPPDHNVHIHEGNIVAMQGEAMITQARST	775		
Db	794	EEVYDWM-----YTQREVHFQKGNKNETLDAPLGIPLHIRGGNVLPTQEPGYTVAESRON	848		

QY	776	PFHLVVMDSHVASTGELFLONGIEMDIGGPGKWTLLVRRFAESGINNLTTSSEVWVRGY	835		
Db	849	PFGLIVALNDGKAGGSLYLDGSLVDSS-----LLVSP-----SVSDNTUSAS-PSGDY	899		
QY	836	AMSQRWMDKXITILGLKRRVK-IX-----EYTVQKDAAGAIKVKGLGRRTSSHNGGF	886		
Db	900	KADQ--PLANVTILGVGHKPKSVKFNANVDFTYKK--STVFVTGLDKYT---KDGAF	950		
RESULT 13					
YAJI_SCHPO					
ID	YAJI_SCHPO	STANDARD;	PRT;	993 AA.	
AC	Q09901;				
DT	01-FEB-1996 (Rel. 33, Created)				
DT	01-OCT-1996 (Rel. 34, Last sequence update)				
DT	28-FEB-2003 (Rel. 41, Last annotation update)				
DE	Putative family 31 glucosidase C30D11.01c precursor (EC 3.2.1.1.-).				
GN	SPAC30D11.01C OR SPAC56F8.01				
OS	Schizosaccharomyces pombe (fission yeast).				
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;				
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;				
OC	Schizosaccharomycetes.				
OX	NCBI_TaxID=4896;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=972;				
RC	MEDLINE=21848401; PubMed=11859360;				
RA	Wood V., William R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,				
RA	Sgourou J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,				
RA	Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,				
RA	Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,				
RA	Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,				
RA	Holroyd S., Hornaby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,				
RA	James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,				
RA	Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,				
RA	Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,				
RA	Rutherford J., Simmonds M., Squares R., Squares S., Stevens K.,				
RA	Skellon J., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,				
RA	Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,				
RA	Welljens I., Vanscreels E., Rieger M., Schaefer M., Mueller-Auer S.,				
RA	Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,				
RA	Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,				
RA	Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,				
RA	Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,				
RA	Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,				
RA	Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,				
RA	Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,				
RA	Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,				
RA	Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,				
RA	Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,				
RT	"The genome sequence of Schizosaccharomyces pombe."				
RT	Nature 415:871-880(2002).				
CC	!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
DR	EMBL; 267961; CAA91887.1; -				
DR	EMBL; 269728; CAA93572.1; -				
DR	PIR; T38598; T38598.				
DR	GeneDB_Spombe; SPAC30D11.01c; -				
DR	InterPro; IPR000322; Glyco hydro_31_1-				
DR	Pfam; PF01055; Glyco hydro_31_1-				
DR	PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; FALSE_NEG.				
DR	PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.				
KW	Hypothetical protein; Hydrolase; Glycosidase; Glycoprotein; Signal.				
FT	POTENTIAL				





CC -!- DISEASE: Defects in SI are the cause of disaccharide intolerance  
 CC I (MIM:222900).  
 CC -!- MISCELLANEOUS: THERE IS A HIGH DEGREE OF HOMOLOGY BETWEEN THE  
 CC ISOMALTASE AND SUCRASE PORTIONS (41 % OF AMINO ACID IDENTITY)  
 CC INDICATING THAT THIS PROTEIN IS EVOLVED BY PARTIAL GENE  
 CC DUPLICATION.  
 CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.  
 CC -!- SIMILARITY: Contains 1 P-type (trefoil) domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EXBL: X63597; CAA45140.1; -;  
 CC EMBL: M22616; AAA60551.1; ALT\_SEQ.  
 CC PIR: S36082; UOHU.  
 CC Genew: HGNC:10856; SI.  
 CC MIM: 222900; -;  
 CC GO: GO:0005903; C:brush border; TAS.  
 CC GO: GO:0005794; C:Golgi apparatus; TAS.  
 CC InterPro: IPR000322; Glyco\_hydro\_31.  
 CC InterPro: IPR000519; P\_trefoil.  
 CC Pfam: PF01055; Glyco\_hydro\_31; 2.  
 CC Pfam: PF00088; trefoil; 2.  
 CC SMART: SW00018; PD; 2.  
 CC PROSITE: PS00025; P\_TREFOIL; 1.  
 CC PROSITE: PS00129; GLYCOSYL\_HYDROL\_F31\_1; 2.  
 CC PROSITE: PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.  
 CC Multifunctional enzyme; Transmembrane; Glycoprotein; Hydrolase;  
 CC Glycosidase; Repeat; Signal-anchor; Sulfation; Disease mutation.  
 CC INIT MET 0 BY SIMILARITY.  
 CC CHAIN 1 1826 SUCRASE-ISOMALTASE, INTESTINAL.  
 CC CHAIN 2 1006 ISOMALTASE.  
 CC DOMAIN 1007 1826 SUCRASE.  
 CC DOMAIN 1 11 CYTOPLASMIC.  
 CC TRANSMEM 12 31 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).  
 CC DOMAIN 32 1826 LUMENAL.  
 CC DOMAIN 42 59 SER/THR-RICH.  
 CC DOMAIN 61 108 P-TYPE.  
 CC DOMAIN 109 1006 ISOMALTASE.  
 CC DOMAIN 1007 1826 SUCRASE.  
 CC ACT\_SITE 504 504  
 CC ACT\_SITE 1393 1393  
 CC DISULFID 62 93 BY SIMILARITY.  
 CC DISULFID 76 92 BY SIMILARITY.  
 CC DISULFID 87 105 BY SIMILARITY.  
 CC MOD\_RES 236 236 SULFATION (POTENTIAL).  
 CC MOD\_RES 238 238 SULFATION (POTENTIAL).  
 CC MOD\_RES 390 390 SULFATION (POTENTIAL).  
 CC MOD\_RES 399 399 SULFATION (POTENTIAL).  
 CC MOD\_RES 666 666 SULFATION (POTENTIAL).  
 CC MOD\_RES 762 762 SULFATION (POTENTIAL).  
 CC MOD\_RES 764 764 SULFATION (POTENTIAL).  
 CC CARBOHYD 98 98 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC CARBOHYD 436 436 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC CARBOHYD 454 454 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC CARBOHYD 660 660 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC CARBOHYD 903 903 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC CARBOHYD 1234 1234 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC CARBOHYD 1302 1302 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC CARBOHYD 1339 1339 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC CARBOHYD 1353 1353 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC CARBOHYD 1402 1402 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC CARBOHYD 1747 1747 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC CARBOHYD 1762 1762 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC VARIANT 1097 1097 Q -> P (in disaccharide intolerance I;  
 CC exhibits intracellular accumulation of  
 CC mannose-rich SI in the Golgi).  
 CC /FTid=VAR\_007854.

FT	CONFLICT	661	661	MISSING (IN REF. 2).
SQ	SEQUENCE	1826 AA;	20927272 MW;	3F7E4B66FDCF9C8E CRC64;
	Query Match	26.8%	Score 1304;	DB 1; Length 1826;
	Best Local Similarity	34.6%	Pred. No. 4.6e-79;	
	Matches 311;	Conservative 157;	Mismatches 316;	Indels 114; Gaps 28;
QY	43	YGYOVNAKVDNSTGKSTALLQIRNSPVVGPDIHFLSFTASPEEDDTLRIFFDANNR	102	
DB	111	HGVNVQDM---TTTSGVEAKLRIPSTLFGNDINSVLFTTQNTNRPFRPKITDPNR	167	
QY	103	RWEIPNEVLPRPPPPPPSLSSLOHLPKIPQNOPTTVLSLPHSDLAFTLFTPTPGF	162	
DB	168	RYEPHQYKVEFTGPTVSDTLVDVK-----VAQN-----PPSI	200	
QY	163	TIYKSTHVDLFDATPISNPETTELKYDOVLQSSLPAAQAAHLYGLGEHTKTFQLAH	222	
DB	201	QVIRKSNKTLTDFTSIGP-----LVYSDQYLQISARLPSD--YIYGIGEOVHKRFR--H	250	
QY	223	NOILLTLWNADIASFNRDL-----NLYGSHPFYMDVRSSPMVSGTHGVFLNSGMDE	275	
DB	251	DLSMKTM---PIFTRDOLPGDNNNNLYGHOTPFMCIEDTS--GKSGVFLMNSNAMEIF	304	
QY	276	YTGDRIT-YKVIIGIIDLIIYIFAGRTPEMWLDQYTKLIGRPAPMPYNAEPGHCRCNGYRDV	334	
DB	305	IQPTPIVTVRTVGTGILDFYILLGDTPEQVVOYQOVLGPAMPAYMNLGQLSRMNYKSL	364	
QY	335	NEIETVVDKYAEARIPLEVMWMTDIDYMDAFKDTLDVPVHFFLDKMOQFVTKLHRNGORYV	394	
DB	365	DVVEVVRNRREAGIPFDTQVTDIDYMDKDFDYDQVAF--NGLPQVQDLHDHGQYV	422	
QY	395	PILDPGINTNK-----SYGTPIRGMQSNVFIKRN--GNPYLGSVMPGPVYVDFLDPAA	447	
DB	423	IILDPAISIGRRANGTTVATYVERGNTQHWLNESDGTPIIGEVMPGLTVVPDFTNPCI	482	
QY	448	SFWDIEIKRFRDILPIDGIWDMNEASNFITSATPG---STLDNPPYKINNSGRVP--	502	
DB	483	DWANECSIFHQEVQYDGLWDMNEVSSFIQGS-TKGCNVNKLNYPPF-----TPDI	533	
QY	503	---INSKTIPTATAM-HYGVNTEYNAHNLGLFLESQATREAL--VRPATRGPELLSRSTF	555	
DB	534	LDKLMYSKTIQMDAVQNMKG--QYDVHSLYGSMAIATEQAVQVFPNKR8-FILTRSTF	590	
QY	556	AGSGKYTAHTWTGDNAARDLQYISPTMFLGFMGPMIGADICGFABSTTEELCCRWIQ	615	
DB	591	AGSGRHAHMLGDNATASWEQWEMSIITGLMFLSGIPLVGADICGFVAETTELCCRWQ	650	
QY	616	LGAFYFPRSDHSARDTTHQBYLYM---ESVAASARTVLGLRYELLPPYYITLMYDANLGS	672	
DB	651	LGAFYFPRSRNHSNGYEHQDPAFFQNSLLVSKSRQYLTIRYTLPLFLYLYKAHVEGE	710	
QY	673	PIARPLSFTFPDDVATYGISQFLIGRIMVSPVLQPGSSIVNAYSPRGNVSLSNYTS	732	
DB	711	TVARPLVHEFYEDTNSWIEDTEFLWGPALLIPLVKKQADTVSAIYDAIYV---DYESG	767	
QY	733	VSVS-AGTVSVLSAPPDINHVIHIEGNIVAMQGEAMTTQAAARSTPFFHLVVMSDHVA	791	
DB	768	AKRPMRKQVDMYLPADKIGLHLRGYIIPQEPDVTITASKRKNPLGLIVALGENTAKG	827	
QY	792	ELFLDNGIEMDIGGPGKWTLVRFPAESGINNLISSEVNRGYSAMSORWMDKITLGL	851	
DB	828	DFFWDDGETKDT-IQNGNYIILYTF-----SVSNNTLIDIVCTHSSYQSGTTLA	882	
QY	852	KRRVKIKYTVQKDAGAIKKVGLGRTRTSSHNGGFFVS---VISDLKQLVGOAFKLE	905	
DB	883	TDSVT-----EVRVAENQPMNAHNFITDASNQVLLIADLKINLGRNFSVQ	929	

Search completed: October 27, 2003, 10:20:27

Job time : 19.7086 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 ; Search time 23.8921 Seconds  
(without alignments)  
3678.964 Million cell updates/sec

Title: US-10-043-418-2

Perfect score: 4862

Sequence: 1 MERSKLPYICPTLAVLPL.....RLVQQAFLKLEFEGATRV 914

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 3

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76:\*

1: piri:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4858	99.9	913	JC5463	alpha-glucosidase
2	3243.5	66.7	903	T09143	alpha-glucosidase
3	2719.5	55.9	902	T48531	alpha-glucosidase
4	2370	48.7	877	S65057	alpha-glucosidase
5	2045	42.1	915	H96709	hypothetical prote
6	1849	38.0	855	T47534	hypothetical prote
7	1460.5	30.0	864	J44624	alpha-glucosidase
8	1450.5	29.8	952	A32609	alpha-glucosidase
9	1355.5	27.9	1827	A23945	sucrose alpha-gluc
10	1329	27.3	995	T50267	probable family 31
11	1324.5	27.2	958	JN0102	glucan 1,4-alpha-g
12	1321.5	27.2	993	T38598	probable family 31
13	1307	26.9	985	JC4217	alpha-glucosidase
14	1304	26.8	1827	UWHU	sucrose alpha-gluc
15	1292.5	26.6	1841	T10799	sucrose alpha-gluc
16	1161.5	23.9	1070	S19886	alpha-glucosidase
17	1138	23.4	1743	T15893	hypothetical prote
18	1090.5	22.4	719	JC1200	alpha-glucosidase
19	1060.5	21.8	919	T16693	hypothetical prote
20	938	19.3	856	T22575	hypothetical prote
21	882.5	18.2	763	AH1097	alpha-glucosidase
22	868.5	17.9	763	AG1460	alpha-glucosidase
23	837	17.2	954	A62105	glucan 1,4-alpha-g
24	828.5	17.0	779	AE2402	alpha-glucosidase
25	816	16.8	919	T07391	probable alpha-glu
26	786.5	16.2	818	AC2472	alpha-glucosidase
27	782	16.1	910	T22050	hypothetical prote
28	778	16.0	924	T22044	hypothetical prote
29	761	15.7	941	T32449	hypothetical prote

30	696	14.3	693	2	H90486	alpha-glucosidase
31	541.5	11.1	642	2	S11386	sucrose alpha-gluc
32	481.5	9.9	764	2	A72394	alpha-xylosidase -
33	479.5	9.9	772	2	B65167	hypothetical 88.1
34	473.5	9.7	772	2	D91195	hypothetical prote
35	473.5	9.7	772	2	B86042	hypothetical prote
36	470	9.7	731	2	D90483	alpha-xylosidase (
37	453	9.3	772	2	A10968	probable glycosyl
38	444.5	9.1	773	2	A83888	hypothetical prote
39	428.5	8.8	983	2	B87347	glycosyl hydrolase
40	424	8.7	801	2	H83737	glucosidase BH0704
41	418	8.6	769	2	H97033	alpha-glucosidase
42	392.5	8.1	792	2	AD0104	probable glucosida
43	373.5	7.7	679	2	AD0507	probable glycosyl
44	368.5	7.6	657	2	G83906	hypothetical prote
45	367	7.5	275	2	A36690	sucrose alpha-gluc

RESULT 1

JC5463

alpha-glucosidase (EC 3.2.1.20) - sugar beet

N:Alternate names: alpha-D-glucoside glucohydrolase

C:Species: Beta vulgaris var. altissima (sugar beet)

C:Date: 17-Jun-1997 #sequence\_revision 12-Sep-1997 #text\_change 20-Jun-2000

C:Accession: JC5463; PC4330

R:Matsui, H.; Iwanami, S.; Ito, H.; Mori, H.; Honma, M.; Chiba, S.

Biosci. Biotechnol. Biochem. 61, 875-880, 1997

A:Title: Cloning and sequencing of a cDNA encoding alpha-glucosidase from sugar beet.

A:Reference number: JC5463; MUID:97321863; PMID:9178565

A:Accession: JC5463

A:Molecule type: mRNA

A:Residues: 1-913 <Matl>

A:Cross-references: DDBJ:D89615; NID:g2217947; PIDN:BAA20343.1; PID:g2190276

A:Experimental source: seed; cv. NK-152

A:Accession: PC4330

A:Molecule type: protein

A:Residues: 234-261;310-365;507-541;810-840 <MAT2>

A:Experimental source: seed

C:Comment: This enzyme is an alpha-glucosidase that catalyzes the hydrolysis of alpha-glucosides.

C:Superfamily: lysosomal alpha-glucosidase; sucrose/isomaltase homology; trefoil homo.

C:Keywords: glycosidase; hydrolase

F:149-803/Domain: sucrose/isomaltase homology <SIM>

Query Match	99.9%;	Score	4858;	DB 2;	Length	913;	
Best Local Similarity	100.0%;	Pred. No. 0;					
Matches	913;	Conservative	0;	Mismatches	0;	Gaps	0;
QY	1	MERSKLPYICPTLAVVLPLVLCMVVEGATTSKNDNGEALGYQVKNKAVDNSTGKSL	60				
Db	1	MERSKLPYICPTLAVVLPLVLCMVVEGATTSKNDNGEALGYQVKNKAVDNSTGKSL	60				
QY	61	TALLQLIRNSPVYGPDIHFLSFTASFEEDDTLRIRFTDANNRWEIPNEVLPRPPPPSP	120				
Db	61	TALLQLIRNSPVYGPDIHFLSFTASFEEDDTLRIRFTDANNRWEIPNEVLPRPPPPSP	120				
QY	121	PPLSSLOHLPKPIQONQPTTVLSHPHSDLAFTLFTHTTPGFTIYKKSHTDVLDFATPIP	180				
Db	121	PPLSSLOHLPKPIQONQPTTVLSHPHSDLAFTLFTHTTPGFTIYKKSHTDVLDFATPIP	180				
QY	181	SNPTFFLIYKDYQLQSSLPQAQAHLYGGEHTKPTFQLAHNOILTLWNADIASFNRLD	240				
Db	181	SNPTFFLIYKDYQLQSSLPQAQAHLYGGEHTKPTFQLAHNOILTLWNADIASFNRLD	240				
QY	241	NLYGSHFFYMDVRSSPMVGSTHGVFLNSNGMDVEYTGDRITTKVIGIIDLIFAGRT	300				
Db	241	NLYGSHFFYMDVRSSPMVGSTHGVFLNSNGMDVEYTGDRITTKVIGIIDLIFAGRT	300				
QY	301	EMVLDDQYTKLIGRPAPMPYNAFGHCQCRGWGRDVNEIETVVDKYAEARIPLEVMWTDIDY	360				
Db	301	EMVLDDQYTKLIGRPAPMPYNAFGHCQCRGWGRDVNEIETVVDKYAEARIPLEVMWTDIDY	360				

Qy 361 MDAFKDTLPVHPLDKMQQVTKLHRNGQRYVPLDPCINTNKSYGTFIRMQSNVFI 420  
Db 361 MDAFKDTLPVHPLDKMQQVTKLHRNGQRYVPLDPCINTNKSYGTFIRMQSNVFI 420  
Qy 421 KRNGNPVLGSWPGVYVPDFLDPAAARSFWVDEIKRFRDILPDIGTWIDMNEASNFITSA 480  
Db 421 KRNGNPVLGSWPGVYVPDFLDPAAARSFWVDEIKRFRDILPDIGTWIDMNEASNFITSA 480  
Qy 481 PTGSLTLDNPPYKINNSGGRVPINSKTIPTAMHYGNVTEYNAHNLVGFLESQATREALV 540  
Db 481 PTGSLTLDNPPYKINNSGGRVPINSKTIPTAMHYGNVTEYNAHNLVGFLESQATREALV 540  
Qy 541 RPATRGPFLLSRSTFAGSGKYTAHWTGDNAAARDLQYSIPTMLNFGFMPIGADICG 600  
Db 541 RPATRGPFLLSRSTFAGSGKYTAHWTGDNAAARDLQYSIPTMLNFGFMPIGADICG 600  
Qy 601 FAESTTEELCCRMVQLGAFYPPFSRDHSARDTHQELYLWESVAASARTVLGLRYELLPPY 660  
Db 601 FAESTTEELCCRMVQLGAFYPPFSRDHSARDTHQELYLWESVAASARTVLGLRYELLPPY 660  
Qy 661 YTLMYDANLRGSPARPLSFTFPDDVATYGISOFLIGRIMVSPVLQPGSSIVNAYSPP 720  
Db 661 YTLMYDANLRGSPARPLSFTFPDDVATYGISOFLIGRIMVSPVLQPGSSIVNAYSPP 720  
Qy 721 GNMVLSNYSVSSVSAGTYVLSAPPDHINVHIHEGNIIVAMOGAEMTTQAAESTPPLL 780  
Db 721 GNMVLSNYSVSSVSAGTYVLSAPPDHINVHIHEGNIIVAMOGAEMTTQAAESTPPLL 780  
Qy 781 VVMSDHSVASTGELFDNGIEMDGGPGKMTLVRFFAESGINNLTISSEVNVNGYAMSQR 840  
Db 781 VVMSDHSVASTGELFDNGIEMDGGPGKMTLVRFFAESGINNLTISSEVNVNGYAMSQR 840  
Qy 841 WMDKTIILGLKRVKIKEYTVQKDAGAIKVGIGRATSSHNOGGFPVSVISDLRLQVG 900  
Db 841 WMDKTIILGLKRVKIKEYTVQKDAGAIKVGIGRATSSHNOGGFPVSVISDLRLQVG 900  
Qy 901 AFKLELEFEAGTR 913  
Db 901 AFKLELEFEAGTR 913  
RESULT 2  
T09143  
alpha-glucosidase (EC 3.2.1.20) - spinach  
C:Species: Spinacia oleracea (spinach)  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: T09143  
R:Sugimoto, M.; Furui, S.; Suzuki, Y.  
Plant Mol. Biol. 33, 765-768, 1997  
A:Title: Molecular cloning and characterization of a cDNA encoding alpha-glucosidase from  
A:Reference number: 216585; MUID:97238484; PMID:9132069  
A:Accession: T09143  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-903 <SUG>  
A:Cross-references: EMBL:D86624; NID:g2081626; PIDN:BAA19924.1; PID:g2081627  
A:Experimental source: strain Dash  
C:Superfamily: lysosomal alpha-glucosidase; sucrose/isomaltase homology; trefoil homology  
C:Keywords: glycosidase; hydrolase  
F:145-799/Domain: sucrose/isomaltase homology <SIM>  
Query Match 66.7%; Score 3243.5; DB 2; Length 903;  
Best Local Similarity 67.3%; Pred. No. 2e-221;  
Matches 610; Conservative 122; Mismatches 166; Indels 9; Gaps 5;  
Qy 3 RSKLPRIYICPTLAVLPLVLCMVVEGATTSKNDQGEAIGYGVQVKNVKNYSTGKSLTA 62  
Db 2 KKKIPSL---ALGILLVFLQYLVAGISTENDPEG-VIGYGVKXSVKVDSDGSTRSLTA 57  
Qy 63 LLOLRKSPVYGPDIHPLSFTASFEEDTLIRITDANNRWEIPNEVLPR-ppppsp 121  
Db 58 LPQLVKNSSVYGPDIQLLSITASLESDNRURVITDAKRRWEIPDNILRHQPPPPPH 117

Qy 122 PLSSQLHLPKPIPONQPTTTVLSPHPSDLAFTLPHHTTFFGFTIYRKSTHVDLFDATPIS 181  
Db 118 SLSSLYRTLSSPTTNRKILLSPNSDLTPSLINTTFFGFTISRKSTHVDLFDATDPT 177  
Qy 182 NPTFLIYKDOYLQSSSLPAQAAHLYGLGEHTKPTFQLAHNOILTLWNADIASFNRLN 241  
Db 178 NPNTFLIFIDYLLHTLSSLPOTRAHIIYGLGEHSKPTFQLAHNOILTLWNADIASFNRLN 237  
Qy 242 LYGSHPPFMDYRSSPMVSGTHGVFLNNSGMDVBYTGDRITKYKVIIGIIDLIIYFAGRTPE 301  
Db 238 LYGSHPPFMDYRSSPMVSGTHGVFLNNSGMDVBYTGDRITKYKVIIGIIDLIIYFAGSPG 297  
Qy 302 MYLDQYTKLGRPAPMYKAFGHQCRWGYDNYEIEIVVDYKAYEARLPLEVMWMTDIDYM 361  
Db 298 QVVEQFTRVIGRPPMPYMAFGQQCRGYDHYVDELOSVVAGYAKAKIPLVEMWMTDIDYM 357  
Qy 362 DAFKDTLPDVFHFLDKMQQVTKLHRNGQRYVPLDPCINTNKSYGTFIRMQSNVFIK 421  
Db 358 DAYKDTLPDVPNFPDLDMKKEFVNLHKNQKYVVLDPGISTNKTYETIRGMKGDVFLK 417  
Qy 422 RGNPVLGSWPGVYVPDFLDPAAARSFWVDEIKRFRDILPDIGTWIDMNEASNFITSA 481  
Db 418 RGNKPYLGSWPGVYVPDFLKPASALTFTWDEIKRFLNLLPVDGLWIDMNEISNFISPP 477  
Qy 482 TPGSLTLDNPPYKINNSGGRVPINSKTIPTAMHYGNVTEYNAHNLVGFLESQATREALV 541  
Db 478 TPGSLTLDNPPYKINNSGGRVPINSKTIPTAMHYGNVTEYNAHNLVGFLESQATREALV 537  
Qy 542 PATRGPFLLSRSTFAGSGKYTAHWTGDNAAARDLQYSIPTMLNFGFMPIGADICG 601  
Db 538 LTERPFLVLSSTFSGSGKYTAHWTGDNAAARDLQYSIPTMLNFGFMPIGADICG 597  
Qy 602 ABSTTEELCCRMVQLGAFYPPFSRDHSARDTHQELYLWESVAASARTVLGLRYELLPPY 661  
Db 598 LGNTTEELCCRMVQLGAFYPPFSRDHSARDTHQELYLWESVAASARTVLGLRYELLPPY 657  
Qy 662 TLMYDANLRGSPARPLSFTFPDDVATYGISOFLIGRIMVSPVLQPGSSIVNAYSPP 721  
Db 658 TLMYDANLRGSPARPLSFTFPDDVATYGISOFLIGRIMVSPVLQPGSSIVNAYSPP 717  
Qy 722 GNMVLSNYSVSSVSAGTYVLSAPPDHINVHIHEGNIIVAMOGAEMTTQAAESTPPLL 781  
Db 718 GNMVLSNYSVSSVSAGTYVLSAPPDHINVHIHEGNIIVAMOGAEMTTQAAESTPPLL 777  
Qy 782 VVMSDHSVASTGELFDNGIEMDGGPGKMTLVRFFAESGINNLTISSEVNVNGYAMSQR 841  
Db 778 VVMSDHSVASTGELFDNGIEMDGGPGKMTLVRFFAESGINNLTISSEVNVNGYAMSQR 837  
Qy 842 WMDKTIILGLKRVKIKEYTVQKDAGAIKVGIGRATSSHNOGGFPVSVISDLRLQVG 899  
Db 838 WMDKTIILGLKRVKIKEYTVQKDAGAIKVGIGRATSSHNOGGFPVSVISDLRLQVG 895  
Qy 900 QAFKLEL 906  
Db 896 REFLEL 902  
RESULT 3  
T48531  
alpha-glucosidase 1 - Arabidopsis thaliana  
N:Alternate names: protein T22P22.110  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 21-Jul-2000  
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancro  
submitted to the Protein Sequence Database, April 2000  
A:Reference number: Z24490  
A:Accession: T48531  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-902 <BEV>  
A:Cross-references: EMBL:AL163814



A:Experimental source: cultivar Columbia; BAC clone T22P22

C:Genetics:

A:Map position: 5

A:Introns: 78/2; 313/1; 390/1; 605/3; 747/1

A:Note: T22P22.110

C:Superfamily: lysosomal alpha-glucosidase; sucrase/isomaltase homology; trefoil homology

Query Match 55.9%; Score 2719.5; DB 2; Length 902;  
Best Local Similarity 57.0%; Pred. No. 2.7e-184;  
Matches 516; Conservative 146; Mismatches 219; Indels 25; Gaps 7;

```
QY 12 PTLAVVLPLVLCMVVEGATTSKNDNGEAGYGYQVNAKVNSTGKSLTALLQIRNSP 71
DB 8 PNIFIV- -VVVFFSLRSSQVLEEEESTVVGYYVRSVGVDP- -SNRQVLTAKLDLIKPS 64
QY 72 VYGPDIHFLSFTASFEEDTLRFRFDANNRWEIPEVLPNEVLPRPPPPPPSLSSLOHLPK 131
DB 65 VTAPDIKSLNLHVSLETSLRIRITDSSQORWEIPEVLP- - - - -AGNHSFR 113
QY 132 PIPQ- - - - -NQPTTTLVLSHPHSDLAFTLPHFTTFFGFTIYRKSTHDLVDFATPIPSNPTTF 186
DB 114 RSTBEDGCGSPENNFLADPSDDLVTFLHTTFFGFSVSRSSGGDILFDTSPDSDSNITY 173
QY 187 LIYKOYLQSSSLPAQOAHLYGLGHEHTPTTOLAHNQILTLWNADIASFNRLNLYGSH 246
DB 174 FIFKOQFLQSSALPENRNLXGIGHEHTKRSFRLIPGETMTLWNADIGSENPDVNLGSH 233
QY 247 PFYMDVRSS- - -PMVSGTHGVFLNSGMDVEYTGDRITYKYVIGGIIDLYIFAGRTPEMV 303
DB 234 PFYMDVRSGKNEEAGTTIGVLLNSGMDVKYEGHRIYINVGIVDLYVFAGPSPEVY 293
QY 304 LDQYTKLIGRPAMPYNAFGHOCRWGYRDVNEIETVVDKYAEARIPLEVMTDIDYMDA 363
DB 294 MNQYTELICRPAMPYNSFGHOCRWGYKNVSDLEVYVDGYAKAGIPLEVMTDIDYMDG 353
QY 364 FXDFTLDVPHFLDKMQQFVTKLHRNGORYVPLDPGINTKNSYGFPIRGMSNVFIKRN 423
DB 354 YKDFTLDPVNFPEDKMQSFVDTLHKNGQYVILDPGIGVDSYGYTNRGMEADVFIKRN 413
QY 424 GNPYLGSMVGPVYVDFLDPAAARSFWDEIKRFDILPDIGLWIDMNEASNFIISAPTP 483
DB 414 GNPYLGSMVGPVYVDFLDPAAARSFWDEIKRFDILPDIGLWIDMNEASNFIISAPTP 473
QY 484 GSTLDNPPYKINNSGGRVPINSKTIPTATMAYGNVTEYNAHNLGYPLESQATREALVRPA 543
DB 474 GSSLDOPPYKINNSGGRVPINSKTIPTATMAYGNVTEYNAHNLGYPLESQATREALVRPA 533
QY 544 TRGPFLLSRSTFAGSKYTAHWTGDNAAARDWDDLOYSIPTMLNFGLEGMPIGADICGPAE 603
DB 534 GRKPFILSRSTFVSSGKYTAHWTGDNAAKWDLAYSPGILNFGLEGI PMVGADICGFSH 593
QY 604 STTEELCCRWIQLGAFYPSRDSHARDTHOELYLWESVAASARTVLGLRYELLPPYYTL 663
DB 594 DTTEELCCRWIQLGAFYPSRDSHARDTHOELYLWESVAASARTVLGLRYELLPPYYTL 653
QY 664 MYDANLRGSPARPLSFTFPDDVATYGISQSFLIGRIMVSPVLQSGSSINVAISPRGNW 723
DB 654 MYEARYSGNPIARPLSFTFPDDVATYGISQSFLIGRIMVSPVLQSGSSINVAISPRGNW 713
QY 724 VLSLNTSSVSAGYVLSLAPDHPHINHHIHEGNTIVAMQGEAMTTOAARSTPFFHLLVYM 783
DB 714 FDLFNYSFAVGGDSGKHVRDLTPADHVNVRHREGSIVAMQGEALTTRDARKTPYQLLVYA 773
QY 784 SDHVASGTGELFLDNGTDMIDGPGGK- -WTLVRFFAESGINNLTISSEVVNRYGAMSQRW 841
DB 774 SRLNLSGELFLDNGENLRMGAGGRDNTLVKRCYVTVGKSVLRSEVVNPEYASKKRW 833
QY 842 VMDKITILGLKRRVKIKEYTVQDAGAKVYKGLGRRTSSHNQGGFFVSV- -ISDLRLVQQ 900
DB 834 SICKVTFVGFENVENKTEVVRTSERLSRPSRLSIKTVSDNDODPRFLSVEVSKLSLVCK 893
QY 901 AFKLEL 906
|::|
```

DB 894 KFEMRL 899

RESULT 4  
S65057

alpha-glucosidase (EC 3.2.1.20) - barley

C:Species: Hordeum vulgare (barley)

C:Date: 20-Jul-1996 #sequence\_revision 27-Feb-1997 #text\_change 22-Jun-1999

C:Accession: S65057; S65058

R:Tibbot, B.K.; Skadsen, R.W.

submitted to the EMBL Data Library, March 1995

A:Description: Molecular cloning and characterization of a gibberellin-inducible alpha-

A:Reference number: S65057

A:Accession: S65057

A:Molecule type: mRNA

A:Residues: 1-877 <TIB>

A:Cross-references: EMBL:U22450; MID:g944900; PID:AAB02985.1; PID:g944901

R:Tibbot, B.K.; Skadsen, R.W.

Plant Mol. Biol. 30, 229-241, 1996

A:Title: Molecular cloning and characterization of a gibberellin-inducible, putative

A:Reference number: S65058; MID:96178863; PMID:8616248

A:Accession: S65058

A>Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 85-115, 'S', 117-121, 123-170, 'D', 172-366, 'G', 368-529, 'H', 531-685, 'G', 687-76-

A:Cross-references: EMBL:U22450

C:Superfamily: lysosomal alpha-glucosidase; sucrase/isomaltase homology; trefoil homo-

C:Keywords: glycosidase; hydrolase

F:125-769/Domain: sucrase/isomaltase homology <SIM>

Query Match 48.7%; Score 2370; DB 2; Length 877;

Best Local Similarity 51.2%; Pred. No. 1.5e-159;

Matches 466; Conservative 147; Mismatches 235; Indels 62; Gaps 15;

QY 16 VVPLVLCMVVEGATTSKNDNGEAGYGYQVNAKVNSTGKSLTALLQIRNS 70

DB 7 LLLCLCLCLFAPRLCSKEEGFLAARTVLAVVMTGALRAEATGGRSSTG----- 58

QY 71 PVYGPDIHFLSFTASFEEDTLRFRFDANNRWEIPEVLPNEVLPRP-----PPSPSPPP 122

DB 59 -----DVQRLAVYASLETDSRLVRITDADHPRWEVQDIIIPRAPGDVLDHAPASSAP 113

QY 123 LSSLOHLKPKIPQNPQTTTTLVLSHPHSDLAFTLPHFTTFFGFTIYRKSTHDLVDFATPIPSN 182

DB 114 LQG-----RVLSPAGSDLVLTV-HASPERFTVSRRTGDTLFTOTAP---- 153

QY 183 PTTFLIYKDYQLQSSSLPAQOAHLYGLGHEHTKTFOLAHNQILTLWNADIASFNRLN 242

DB 154 ---GLVFRDKYLEVTSALPAGRLYGLGHEHTKSSFLRHNDSTFLWNADIGASYVDNL 210

QY 243 YGSHPFYMDVRSSPMVSGTHGVFLNSGMDVEYTGDRITYKYVIGGIIDLYIFAGRTPEM 302

DB 211 YGSHPFYMDVRAP---GTAHGVLLSSNGMDVLYGGSVYTVKVLGGVLDFFVFAFPPLA 267

QY 303 VLDQYTKLIGRPAMPYNAFGHOCRWGYRDVNEIETVVDKYAEARIPLEVMTDIDYMD 362

DB 268 VVDQYTLIARAPAMPYNSFGHOCRWGYKNVSDLEVYVDGYAKAGIPLEVMTDIDYMD 327

QY 363 AFKDFTLDPVHFLDKMQQFVTKLHRNGORYVPLDPGINT- - -NKSYGTFIRMQSNVF 419

DB 328 GFKDFTLDRVNFATAELRPFVDRLHRNAQKLYLIDFGIRVDPIDATYGTFTVRMQQDIF 387

QY 420 IKRGNPNVLSGSMVGPVYVDFLDPAAARSFWDEIKRFDILPDIGLWIDMNEASNFIIS 479

DB 388 LKRGNTFVGNVWGDVYFPDFMHPAAAEFWAREISLFRRTIPVDGLWIDMNEISNFVN- 446

QY 480 APTPGSTLDNPPYKINNSGGRVPINSKTIPTATMAYGNVTEYNAHNLGYPLESQATREAL 539

DB 447 -PEPMNALDDPPYRINNDGTGRPINNKTVRLAVHYGVTEYEENLFGLEARATGRV 505

QY 540 VRPATRGPFLLSRSTFAGSKYTAHWTGDNAAARDWDDLOYSIPTMLNFGLEGMPIGADIC 599

DB 506 LRDTRGRPPFLLSRSTFVSGRYTAYTGDNAATWGLRLYSINTMLSFGLFGMPMIGADIC 565

QY	600	GFAESTTEELCCRWIQOLGAFYPSRDSARDTHQELYLWESVAASARTVLGLRYELLPY	659
Db	566	GFNGNTEELCCRWIQOLGAFYPSRDSHSAFTVRRELYLWPSVAASGRKALGLRYQLLPY	625
QY	660	YYTLMDANLRGSPARPLSFTFPDDVATYGISQFLIGRIMVSPVLOPGSSIVNAYS	719
Db	626	FYTLMEAHMTGAPIARPLFSYPHDVATYGVDRQFLGRLVSPVLEPGPTTVDAYFP	685
QY	720	RGNWVSUSNTSSVSAGTYVLSAPPDHNHIEGNIIVANQGBAMTTOAARSTPFSU	779
Db	686	AGRWYRLDYSLAVATRTGKHVRLPAPADTVNHLTGTTILPQOSALATSRARRTAFEL	745
QY	780	LVMSDHVASTGELFLDNGIEMDIGGPGGKWTLLVRFAESGIN--LTISSEVWNGYAM	837
Db	746	LVALAEDGTASGFLFLDDGDSPEY-GRSDMSVRFNYKIPNNKGAIKVKSEVHNSYAO	804
QY	838	SQRMVMDKITILGLKRVKIKKEYTVQKDAGAIKVKGLGRETSSHNOGGF-FVSVISDLRQ	896
Db	805	SRTLIVISKVVLGHRSPAPFKLTVHNSAEVASS-SAGTRYONAGGLGVAAHIGLSL	863
QY	897	LVQGAFKLEL 906	
Db	864	WGEEPELKV 873	
RESULT 5			
H96709			
hypothetical protein F24J5.20 [imported] - Arabidopsis thaliana			
C:Species: Arabidopsis thaliana (mouse-ear cress)			
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001			
C:Accession: H96709			
R:Theologas, A.; Ecker, J.R.; Palm, C.J.; Federici, N.A.; Kaul, S.; White, O.; Alonso, J.;			
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;			
anssen, N.F.; Hughes, B.; Huizar, L.			
Nature 408, 816-820, 2000			
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.			
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,			
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.			
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,			
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.			
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.			
A:Reference number: A86141; MUID:21016719; PMID:11130712			
A:Accession: H96709			
A:Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-915 <STO>			
A:Cross-references: GB:AE005173; NID:G5734722; PIDN:AA49987.1; GSPDB:GN00141			
C:Genetics:			
A:Gene: F24J5.20			
A:Map position: 1			
Query Match 42.1%; Score 2045; DB 2; Length 915;			
Best Local Similarity 43.8%; Pred. No. 1.7e-136;			
Matches 409; Conservative 166; Mismatches 282; Indels 76; Gaps 16;			
QY	13	TLAVLPLVLVMEVGATTSKNDNQGAIGYGVQVNAKVDNSTGKSLTALLQIRNSPV	72
Db	11	SLSLLLALILCF---STQSY-----KTIQGYRL--VSIIESPDGGTGYLQVQKQNKI	60
QY	73	YGPDIHLSFTASFEEDTLIRIRFTDANNRWEIPNEVLPRPPPPPPSLSLQLPKP	132
Db	61	YGSDDITLRLVFKHETSLRLRVHITDAKQRMVEPVNLLPREQPPQGVIGSRKSP--	118
QY	133	IPQNPQPTTVLSHPHSDLATLFTHTPFGFTIYKRSKTHDVLFDATPIPSNPTFLYKQ	192
Db	119	-----ITVQIESSGSELIFS-YTDPFTFAVKRRSNHETLF-----NTTSLVFKDQ	163
QY	193	YLQLSLSPQAQAHYGLGHTKPT-FQLAHNQILLTNADIASFNRDLNLGSHFPFYWD	251
Db	164	YLEISTSLP-KEASLYGLGNSQANGIKLVNPEFYTLTDEVSAINLNTDLYGSHPMYWD	222
QY	252	VRSSPMVGSTHGFLNLSNGMDEVYTGDIRITYKVGIGIIDLTYFAGRTPEWLDQVTKLI	311

Db	223	LRNVGCKAYAHVALLLSNGMDVYRGDSLTYKVIIGVDFYFIAGPSPFLNVQYDTQLI	282
QY	312	GRPAMPYWFAGFHQCRWGYRDVNEIETVVDKYAARIPLEVMWMTDIDYMDAFKDFTLDP	371
Db	283	CRPAMPYWSLGFHQCRWGYHNLVSVEDVDVNYKAKIPLDVIWDDHMDGHDFTLNP	342
QY	372	VHPPLDKMQQFVTKLHRCQORYVILDPGINTNKSXGTFIRGMSQSNVPIKRGNGPYLGSV	431
Db	343	VAYPRAKLLAEFLDKHKIKGMKYIIVNDPGIGVNASYGTFOQAAADVPFIKYEGKPFQAQV	402
QY	432	WPGPVYYPDFDLPAARSEFWDEIKRFRDILPIDIGIWMIDNEASNFIT-----S	479
Db	403	WPGPVYFPDFLNPKTVMWGDEIKRFDLVPIDGLWIDMNEVSNFCSLGLCTIPGKQCP	462
QY	480	APTEG-----STLDNPPYKINNNGRVPINSKTIPATAMHYGNVTEYNAHNL	527
Db	463	GEGPGWVCCLDCKNITKTRWDDPPYKINATGWAPVGFKTIATSATHYNGVREYDAHSIY	522
QY	528	GFLESQATREALVRPATRGPELLSRSTFAGSGKTAHWTGDNARWDDLOYSIPTMLNFG	587
Db	523	GFSETIATHKGLNVQGRPEILSRSTFVSGGQYAAHWTGDNQGTWQSLQVSIPTMLNFG	582
QY	588	LFGMPMIGADICGFAESTTEELCCRWIQOLGAFYPSRDSHARDTHOELYLWESVAASAR	647
Db	583	IFGVPWGSIDICGFYPOPTTEELCNRWIEVGAFYFSRDHANYSPRQELIYQDIVADSAR	642
QY	648	TVLGLRYELLPYYTLMYDANLRGSPARPLSFTFPDDVATYGISQFLIGRIMVSPVL	707
Db	643	NALGMRYKILPFLVTLNVEAHMTGAPIARPLFFSFPFTECYGNSRQFLGSSSEFMISPV	702
QY	708	OPGSSIVNAYSPRGNWYSLNSYTSVSAGTYVLSAPPDHNHIEGNIIVANQGBAM	767
Db	703	EOGKTEVALPFPSPGSHYHFMFTQAVSVKNGKRVTLPAFLNVNHLTYONTILPTQGG	762
QY	768	TTQARSTPFFHLVVM-----SDHVASTGELFLDNG--IEMDIGGPGKWTLLVRFAESGIN	822
Db	763	LSKQARTTFFSLVIAFAFAGASEGATGKLYLDEDELPEMKLG--NGQSTYVDFYASVNG	820
QY	823	NLTISSEVWNGYAMQORWMDKITILGLKRVKIKKEYTVQKDAGAIKVK-----GL	874
Db	821	TKMWSQVKEGFKALSKGWIEKVSLGLRGAGVQSEIQINGSWMTKKIEVSKHEHTYVI	880
QY	875	GRTTSSHNQGGFFVSVISDLR--OLVQGAFLK 904	
Db	881	GLEDEENK-----SVMVEVRGLEMLVGKDFNM 908	
RESULT 6			
T47534			
hypothetical protein F16L2.150 - Arabidopsis thaliana			
C:Species: Arabidopsis thaliana (mouse-ear cress)			
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000			
C:Accession: T47534			
R:Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unseid, M.; Mewes, H.W.; Rudd, S.			
submitted to the Protein Sequence Database, March 2000			
A:Reference number: Z24468			
A:Accession: T47534			
A:Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-855 <JOR>			
A:Cross-references: EMBL:AL162459			
A:Experimental source: cultivar Columbia; BAC clone F16L2			
C:Genetics:			
A:Map position: 3			
A:Introns: 69/3; 291/1; 427/3; 699/1			
A:Note: F16L2.150			
C:Superfamily: lysosomal alpha-glucosidase; sucrase/isomaltase homology; trefoil homo			
Query Match 38.0%; Score 1849; DB 2; Length 855;			
Best Local Similarity 41.8%; Pred. No. 1.2e-122;			
Matches 378; Conservative 165; Mismatches 274; Indels 88; Gaps 16;			





Db 719 AGETVARPLLEFPKDSSTWTDHQLLWGEALITVLQAGKAENVTYFPLGLTWYDL---- 775  
 QY 730 TSSVSVA-----GTYVLSAPPDINHVIHEGNVAMOGAMTQ 770  
 Db 776 -QTVPEALGSLPPPPAARPAIHSEGQWVLPADLTINVHLRGYIIPQGGPLITT 834  
 QY 771 AARSTPFLHVLVMSDRVASTGELFLDNGEMDGGGKWTLLVRFFAEGSINNLTISSEV 830  
 Db 835 ESROQPMALAVALTCKGEARGELFMDGSLV-LEGGAYTVQIFLA-----RNNTIVNEL 889  
 QY 831 V----NRGYAMSORVMDKTIILG-----LKRRIKYEYVQKDAKAIKV 871  
 Db 890 VRVTSAGLQ----LQKTVLGVATAPQOVLNGVPSNFTYSPDKVLDI 937

RESULT 9

A23945  
 sucrose alpha-glucosidase (EC 3.2.1.48) / oligo-1,6-glucosidase (EC 3.2.1.10) - rabbit  
 N:Alternate names: small intestinal sucrose/isomaltase (SI)  
 C:Species: Oryctolagus cuniculus (domestic rabbit)  
 C:Date: 03-Mar-1994 #sequence revision 03-Mar-1994 #text\_change 24-May-1996  
 C:Accession: A23945; B25987; A29163  
 R:Huniker, W.; Spiess, M.; Semenza, G.; Lodish, H.F.  
 Cell 46, 227-234, 1986

A:Title: The sucrose-isomaltase complex: primary structure, membrane-orientation, and evolution  
 A:Reference number: A23945; MUID:86245068; PMID:3755079  
 A:Accession: A23945  
 A:Molecule type: mRNA  
 A:Residues: 1-1827<HUN>  
 R:Joestrom, H.; Noren, O.; Christiansen, L.A.; Wacker, H.; Spiess, M.; Bigler-Meier, B.  
 FEBS Lett. 148, 321-325, 1982  
 A:Title: N-terminal sequences of pig intestinal sucrose-isomaltase and pro-sucrose-isomaltase  
 A:Reference number: A25987; MUID:83105704; PMID:7152027  
 A:Accession: B25987  
 A:Molecule type: protein  
 A:Residues: 2-32, 'XXX', '36-38, '1008, 'N', '1010-1014, 'E', 'SJ2>  
 R:Frank, G.; Brunner, J.; Jauser, H.; Wacker, H.; Semenza, G.; Zuber, H.  
 FEBS Lett. 96, 183-188, 1978  
 A:Title: The hydrophobic anchor of small-intestinal sucrose-isomaltase. N-terminal sequence  
 A:Reference number: A29163; MUID:79086207; PMID:729784  
 A:Accession: A29163  
 A:Molecule type: protein  
 A:Residues: 2, 'VNA', '6-32, 'XXX', '36-38 <FRA>  
 C:Comment: Carbohydrate analysis of the mature enzyme complex indicates both N- and O-linked  
 C:Complex: the two product chains remain associated after cleavage  
 C:Function: <ISO>  
 A:Description: oligo-1,6-glucosidase catalyzes the hydrolysis of 1,6-alpha-D-glucosidic  
 A:Pathway: carbohydrate digestion  
 C:Function: <SUC>  
 A:Description: sucrose alpha-glucosidase catalyzes the hydrolysis of the 1,2-alpha-D-glucopyranosidic  
 A:Pathway: carbohydrate digestion  
 C:Superfamily: sucrose/isomaltase; sucrose/isomaltase homology; trefoil homology  
 C:Keywords: carbohydrate digestion; duplication; glycoprotein; glycosidase; hydrolase; m  
 F:2-1007/Product: oligo-1,6-glucosidase (isomaltase chain) #status experimental <ISO>  
 F:13-32/Domain: membrane associated #status predicted <TM>  
 F:43-65/Region: serine/threonine-rich  
 F:63-109/Domain: trefoil homology <TRF1>  
 F:189-840/Domain: sucrose/isomaltase homology <SIM>  
 F:931-977/Domain: trefoil homology <TRF2>  
 F:1008-1827/Product: sucrose alpha-glucosidase (sucrose chain) #status experimental <SUC>  
 F:1062-1734/Domain: sucrose/isomaltase homology <SIM2>  
 F:12/Binding site: carbohydrate (Thr) (covalent) #status experimental  
 F:99,455,859,896,904,1235,1303,1325,1340,1354,1368,1403,1535,1572,1748,1763,1799/Binding  
 F:1007-1008/Cleavage site: Arg-Ile (tryptin) #status predicted

Query Match 27.9%; Score 135.5; DB 1; Length 1827;  
 Best Local Similarity 35.1%; Pred. No. 3.8e-87;  
 Matches 320; Conservative 157; Mismatches 313; Indels 121; Gaps 28;

QY 43 YGYQVKNKVDNSTGSLTALLQILRNSPVYGPDIHFLSTASFEEDDTLRFTDANKR 102  
 Db 112 HGYNVGQM---TTTGTGLRLNRKSTPTFLFGDINNVLTTTESQANLRFLKLDPNKK 168

QY 103 RWEIPNEVLP PPPPPPPSLSSQLHLPKPIPOKQPTTTVLSHPHSDLAFTLFTHTPEGF 162  
 Db 169 RYEVPHQVTFEAGPAATEILYDVQ-----VTEN-----PFISI 201  
 QY 163 TIYKSTHDLVLDATPIPSNPTTFLIYKQYQLQSSSLPAQQAHLGLGHTKTFOLAH 222  
 Db 202 KVIKSNRNLILFDSIGP-----LVYSDQYQLQISTRLPSE--YMYGFEVHKRFR--H 251  
 QY 223 NQILTLNADIASFNRL-----NLYGSHPFYMDVRSSPMVSGTHGVLLNSNGDVE 275  
 Db 252 DLYKWTW-----PIFTRDQHTDDNNNNLYGHQTFPNCIEDT--TKSGFVFLMNSNAMEIF 305  
 QY 276 YTGDRIT-YKVIIGGIIDLYIFAGRTPEMVLQDKYKLGIRPAMPYMAFGHQCRGWYRDV 334  
 Db 306 IQPTPIVTVRVIIGGILDYIFLGDTPQGVVQYQDELIGRPAMPAYMSLQLSRWYNSL 365  
 QY 335 NEIETVDKVAARIPIEVMWMTDIDYMDAFKDFTLDPVHFPLDKMQQFVTKLHNRGORYV 394  
 Db 366 DVKVEVVRNRREALIPFDQVSDIDYMEDKKDFTYDRVAY--NGLPDPFVQDLHDHGQKYV 423  
 QY 395 PILDPGINTNK-----SVGTFFIRGMQSNVFIKRG--NPVLGSMVPGPVVYVDFLDPAAR 447  
 Db 424 IILDPAISINRRASGEAYESTDRGNAQNVWVNESDGTPIVGEVWPGDTVYPTDTSNCI 483  
 QY 448 SFWDEIKRFRDILPIDGIWIDMNEASNFI--TSAPTPGSTLDNPPYKINNNGRVP--- 502  
 Db 484 EWMANEENIFHQEVNDYDGLWIDMNEVSVFQGSNKGCNDNTLNYPY-----IPDIV 535  
 QY 503 ---INSKTIPTAMHYGNVTYNAHNLVFLGSEATREALVR--PATRGPPLLSRSTFAG 557  
 Db 536 DKLMYSKTLCDMSVQYWG-KOYDVHSLYGYSMATATRAVERVFPNKR-FILTRSTFAG 593  
 QY 558 SGKYTAHWTDGNAARWDDQYSIPTMLNFGLGPMIGADICGFAESTTEELCCRWQLG 617  
 Db 594 SGRAAHWLGDTATWEGMEWSITGMLEFGLFGNPLVGADICGFAETTEELCCRWQLG 653  
 QY 618 AFYFSPSRDHSARDTHQELYLW---ESVAASARTVLRLYELLPPYYTLMYDANLRGSI 674  
 Db 654 AFYFSPNRHNDAGPEHQDPAPFGQDSLKVKSSRHYLNIRYTLPLFLYTLFYKAHAFETV 713  
 QY 675 ARPLSFTFPDDVATYIGISSQPLIGRGIMVSPVLPQGSINVAYSPRGNWVLSNYTSVS 734  
 Db 714 ARPVLHEFYEDTNSWEDREFLWGPALLITPVTQGAETVSAYIPDAVWYDYE--TGAKR 771  
 QY 735 VSAGTYSLSPAPPHINVHIHEGIVAMOGAMTTOAARSTPFLHLLVMSDHSVASTGELF 794  
 Db 772 PMRKQREMSLPADKIGLHLRGYIIPQPAVITATSRMPLGLIATNDNDNTAVGDFF 831  
 QY 795 LDNGIEMDIGGGKWTLLVRFAESGINNLTISSEVNVNRGYAMSORVWMDKITILGLKRR 854  
 Db 832 WDDGETKDT-VQNDNYILYTFVAVSNNNLITCTHELISEGTTLA---FQTIKILGVTE 886  
 QY 855 VKIKEYTVQKDAKAIKVKGLGRRTSSH-----NQGGFVSVISDLRQLVGOAFKLE- 905  
 Db 887 V--TQVTVAE-----NQSMSTHSNFTYDPSNQ-----VLLIENFLNLRGFRVQW 931  
 QY 906 ---LEFEGAT 912  
 Db 932 DQTFLESEKIT 942

RESULT 10

T50267  
 probable family 31 glucosidase [imported] - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 21-Jul-2000  
 C:Accession: T50267  
 R:Hunt, C.; Aves, S.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.  
 submitted to the EMBL Data Library, December 1999  
 A:Reference number: Z25031  
 A:Accession: T50267  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA

A;Residues: 1-995 &lt;HUN&gt;

A;Cross-references: EMBL:AL133522; PIDN:CAB63549.1; GSPDB:GN00066; SPDB:SPAC922.02c

A;Experimental source: strain 972h(-); cosmid c922

C;Genetics:

A;Gene: SPAC1039.11c; SPDB:SPAC922.02c

A;Map position: 1

C;Superfamily: Schwanniomyces glucan 1,4-alpha-glucosidase GAM1; sucrose/isomaltase hom

Query Match 27.3%; Score 1329; DB 2; Length 995;

Best Local Similarity 32.0%; Pred. No. 1.1e-85;

Matches 313; Conservative 164; Mismatches 307; Indels 194; Gaps 32;

K;Alternate names: acid maltase; glucoamylase

N;Species: Schwanniomyces occidentalis

C;Date: 10-Mar-1994 #sequence\_revision 05-Apr-1995 #text\_change 18-Jun-1999

C;Accession: JN0102

R;Dohmen, R.J.; Strasser, A.W.M.; Dahlems, U.M.; Hollenberg, C.P.

Gene 95, 111-121, 1990

A;Title: Cloning of the Schwanniomyces occidentalis glucoamylase gene (GAM1) and its

A;Reference number: JN0102; MUID:91071592; PMID:1979298

A;Accession: JN0102

A;Molecule type: mRNA

A;Residues: 1-958 &lt;DOH&gt;

A;Cross-references: GB:M60207; GB:M34666; NID:g169845; PIDN:AAA33923.1; PID:g169846

C;Genetics:

A;Gene: GAM1

C;Superfamily: Schwanniomyces glucan 1,4-alpha-glucosidase GAM1; sucrose/isomaltase h

C;Keywords: glycoprotein; glycosidase; hydrolase; polysaccharide degradation

F;1-2/Domain: signal sequence #status predicted &lt;SIG&gt;

F;13-958/Product: glucan 1,4-alpha-glucosidase #status predicted &lt;MAT&gt;

F;156-876/Domain: sucrose/isomaltase homology &lt;SIM&gt;

F;61,78,197,403,416,513,580,602,813,907/Binding site: carbohydrate (Asn) (covalent) #

F;470/Active site: Asp #status predicted

Query Match 27.2%; Score 1324.5; DB 1; Length 958;

Best Local Similarity 33.2%; Pred. No. 2.1e-85;

Matches 318; Conservative 152; Mismatches 315; Indels 173; Gaps 32;

C;Date: 10-Mar-1994 #sequence\_revision 05-Apr-1995 #text\_change 18-Jun-1999

C;Accession: JN0102

R;Dohmen, R.J.; Strasser, A.W.M.; Dahlems, U.M.; Hollenberg, C.P.

Gene 95, 111-121, 1990

A;Title: Cloning of the Schwanniomyces occidentalis glucoamylase gene (GAM1) and its

A;Reference number: JN0102; MUID:91071592; PMID:1979298

A;Accession: JN0102

A;Molecule type: mRNA

A;Residues: 1-958 &lt;DOH&gt;

A;Cross-references: GB:M60207; GB:M34666; NID:g169845; PIDN:AAA33923.1; PID:g169846

C;Genetics:

A;Gene: GAM1

C;Superfamily: Schwanniomyces glucan 1,4-alpha-glucosidase GAM1; sucrose/isomaltase h

C;Keywords: glycoprotein; glycosidase; hydrolase; polysaccharide degradation

F;1-2/Domain: signal sequence #status predicted &lt;SIG&gt;

F;13-958/Product: glucan 1,4-alpha-glucosidase #status predicted &lt;MAT&gt;

F;156-876/Domain: sucrose/isomaltase homology &lt;SIM&gt;

F;61,78,197,403,416,513,580,602,813,907/Binding site: carbohydrate (Asn) (covalent) #

F;470/Active site: Asp #status predicted

Query Match 27.2%; Score 1324.5; DB 1; Length 958;

Best Local Similarity 33.2%; Pred. No. 2.1e-85;

Matches 318; Conservative 152; Mismatches 315; Indels 173; Gaps 32;

C;Date: 10-Mar-1994 #sequence\_revision 05-Apr-1995 #text\_change 18-Jun-1999

C;Accession: JN0102

R;Dohmen, R.J.; Strasser, A.W.M.; Dahlems, U.M.; Hollenberg, C.P.

Gene 95, 111-121, 1990

A;Title: Cloning of the Schwanniomyces occidentalis glucoamylase gene (GAM1) and its

A;Reference number: JN0102; MUID:91071592; PMID:1979298

A;Accession: JN0102

A;Molecule type: mRNA

A;Residues: 1-958 &lt;DOH&gt;

A;Cross-references: GB:M60207; GB:M34666; NID:g169845; PIDN:AAA33923.1; PID:g169846

C;Genetics:

A;Gene: GAM1

C;Superfamily: Schwanniomyces glucan 1,4-alpha-glucosidase GAM1; sucrose/isomaltase h

C;Keywords: glycoprotein; glycosidase; hydrolase; polysaccharide degradation

F;1-2/Domain: signal sequence #status predicted &lt;SIG&gt;

F;13-958/Product: glucan 1,4-alpha-glucosidase #status predicted &lt;MAT&gt;

F;156-876/Domain: sucrose/isomaltase homology &lt;SIM&gt;

F;61,78,197,403,416,513,580,602,813,907/Binding site: carbohydrate (Asn) (covalent) #

F;470/Active site: Asp #status predicted

```

QY 489 ----NPPYKINSGRVPINSKTIPATAMHYGNVTEYNAHNLGYFLESQATREAL--VRPA 543
    |||||
Db 555 NINYPYAIYVMQSDSLATHAVSPNATHADCTVEVDIHNLYGYLGENATYHALLEFFN 614
QY 544 TRQPELLSRSTFAGSGKYTAHTWGDNAARWDDLOYISPTMLNFGFLGPMIGADICGFAE 603
    |||||
Db 615 KR-PFMISRSFTFRAGKWTGHWGCDNTADWAYAFSIPQAFSPMGIAGLPFFGADVCGFNG 673
QY 604 STTEELCCRMIOICAEYPPSPDSHARDTHOELLYLWESVAASARTVLRLRYELLPPYYTL 663
    ::|||
Db 674 NSOSELCSRMWQSGSFPFFVRNHNLYGAIODEPYWESVAEARTSMALRYLLPPYYTL 733
QY 664 MYDANLRGSPRIARPLSFTFPDDVATYGISSQFLIGRIMVSPVLQPGSSIVNAYSPRG-- 721
    ::|||
Db 734 LHESHTTGLPIILRAFSWQFFNDRSLSGVQNFVGDGLVTVPLEPGVDKVKGVFPFGACK 793
QY 722 -----NWVSLSNITS-SVSVSAGTYVLSAPPDHINVHIHEGNIVAMQGEAMTTOAART 775
    :|||
Db 794 EEVYYDW-----YTQREVHFKDGKNETLDAPLGHIPLIRGGNVLPQTQSPGYTVAESRON 848
QY 776 PFLLVVMSSHVASTGELFLDNGIEMDGGPGKWTILVRFPAESGINNLTSISSEVVRGY 835
    |||||
Db 849 PFLIVALDNDGKAQSGLYLDDGESLVDS-----LLVSP-----SVSDNLTLSAS-PSGDY 899
QY 836 AMSQRWMDKITILGKRRYK-IK-----EYTVQKDAGAIAKYKGLGRRTSSHNQGGF 886
    |||||
Db 900 KADQ--PLANVTILGVGHKPKSVKFENANVDFTYKK--STVFVTGLDKVT---KDGAF 950

RESULT 12
T38598
probable family 31 glycosyl hydrolase (alpha glucosidase) precursor - fission yeast (Sch
C:Species: Schizosaccharomyces pombe
C:Date: 20-Oct-2000 #sequence revision 20-Oct-2000 #text_change 20-Oct-2000
C:Accession: T38598; S62559; T38911
R:Pearson, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, November 1995
A:Reference number: Z21801
A:Accession: T38598
A:Molecule type: DNA
A:Residues: 1-385 <PEA>
A:Cross-references: EMBL:Z67961; PIDN:CAA91887.1; GSPDB:GN000066; SPDB:SPAC30D11.01c
A:Experimental source: strain 972h-; cosmid c30D11
R:Pearson, D.; Churcher, C.M.
submitted to the EMBL Data Library, November 1995
A:Reference number: S62559
A:Accession: S62559
A:Molecule type: DNA
A:Residues: 1-384 <PE2>
A:Cross-references: EMBL:Z67961
R:Pearson, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z21817
A:Accession: T38911
A:Molecule type: DNA
A:Residues: 352-993 <PE3>
A:Cross-references: EMBL:Z69728; PIDN:CAA93572.1; GSPDB:GN000066; SPDB:SPAC56F8.01
A:Experimental source: strain 972h-; cosmid c56F8
C:Genetics:
A:Gene: SPDB:SPAC30D11.01c; SPDB:SPAC56F8.01
A:Map position: 1L
C:Superfamily: Schwanniomyces glaucus 1,4-alpha-glucosidase GAM1; sucrose/isomaltase hom
F:165-384/Domain: sucrose/isomaltase homology (fragment) <SIN>

Query Match 27 28; Score 1321.5; DB 2; Length 993;
Best Local Similarity 32.74; Pred. No. 3.6e-85;
Matches 300; Conservative 167; Mismatches 287; Indels 163; Gaps 27;
QY 44 GYQVKNKAVDNSTGKSLTALLQIRNS-PVYGGPDHFLSFTASFEEDDTLRIRFTDANNR 102
    |||||
Db 84 GYQARNIS-EVSYG--VLAILELAGDCACTAGTDYFLLNLSYDTEERVHISIDNQT 140
QY 103 RWEIPN--EVLPRPPPPPPPLSSQLHLPKPIQONQPTTVLSPHSDIAFTL-----F 155

```

```

    |||||
Db 141 QFQLSNRDWDAP-----LFRSNGFSGNLQYNFSP 172
QY 156 HTTPFGFTYRKSTHVDLFDATIPSNPTTFLIYKQDYLSSSLPAQAAHLYGLGHTK 215
    |||||
Db 173 NTDPFEFWITRIADQVDFDTR--GNP--LIPEDQVIELTNN-VEDYNNVGL-SGSG 224
QY 216 PTFOLAHNQILTLANADIASFNRLNLYGSHPPFVMDVSRSSPMVGST-----HGVFLN 268
    ::|||
Db 225 QSFRLGNLTKTFW-ATGYSDSPEANMYGSHPTTHEORYIP-IGTTNTYTSASHGVLMLS 282
QY 269 SNGMDVEYTGRIYKYVIGGIIDLIYIPAGRT--PEMVLDOYTKLIGRPAMPYWAFFHQ 326
    |||||
Db 283 SNGMEVLLRSYIYKYMIGGIIDLFVYSGSVSPKYTIQQYVQSIGTPTQPYWSLGFQM 342
QY 327 CRWYRDVNELETVDKVAEARIPIEVMTDIDMDAFKPTLDVHPHFDLDMKQOFTKL 386
    |||||
Db 343 SRWGYKTLSDINMRSYLNASNIPTGFMNDIDYMSBFTFTVNSTAPPNPQTLDFRSL 402
QY 387 HRNGQRYVPILDPGI---NTNKS---YGTFRMQSNVPIKR-NGNPYLGSVMPGPVY 438
    |||||
Db 403 DESHQHYVPLDPAIYAANPNKSADRTYYPYSGFEDNIFIKPNGSAYVGMAMPGFVY 462
QY 439 PDFLDPAAARSFWDEI-----KRFRDILPIDGIWDMNEASNF----- 476
    |||||
Db 463 PDFTPAVLQWKQGIILNLSTAFGSNYSDLPFSGCLDMNEPTSPICIGSCGDLKLNP 522
QY 477 -----ITSAP-----TPGS----- 485
    |||||
Db 523 VHPFSLFGVDVNDKSYYPEDFNATNTTEYKSVSRASQOYKATATSEKSHETSSSLSI 582
QY 486 -----TLDNPPYKINSGRVPINSKTIPATAMHYGNVTEYNAHNLGYFLESQATREAL 539
    |||||
Db 583 NGKPEPSINYPYALDSTDTEHDLAQFGVSPNATMHGNTLRLYNTFYGYSEKISFEAL 642
QY 540 --VRPATRGPPLLSRSTFAGSGKYTAHTWGDNAARWDDLOYISPTMLNFGFLGPMIGAD 597
    |||||
Db 643 NSIQPNIR-PFLLSRSTFVSGRYAAHVLGDNKSQSDMVSSISILTNLLGIPMVGAD 701
QY 598 ICGFAESTTEELCCRMIOICAEYPPSPDSHARDTHOELLYLWESVAASARTVLRLRYELL 657
    |||||
Db 702 VCGYNGNTDEELCARWMAIAPLPIFYRNHNSLGIPOEPFRWASVAEASAIIRISLL 761
QY 658 PYTYTLMYDANLRGSPRIARPLSFTFPDDVATYGISSQFLIGRIMVSPVLQPGSSIVNAY 717
    |||||
Db 762 PYWYTLMTASVDGTPWVRPLFFEPFKQISLASVDKQFMIGTALLISPALEPNITYIQGI 821
QY 718 SPRGN---WVSLSNYTSVSVSAGTYVLSAPPDHINVHIHEGNIVAMQGEAMTTOAARS 774
    |||||
Db 822 IPGONDITWYDWTNH-SVINHDYDENITMSAPLGVNIAVRGGNIIPLQOPGYTTYESRN 880
QY 775 TPFHLLVVMSSHVASTGELFLDNGIEMDGGPGKWTILVRFPAESGINNLTSISSEVVRG 834
    |||||
Db 881 NPYSLLIAMDNNGFASGLYIDDGISMQTNSS-----LSVKLNSNSNTITCVVSGTMSVSP 936
QY 835 YAMSQRWMDKITILGL 851
    |||||
Db 937 -----SLANITILGL 946

RESULT 13
JC4217
alpha-glucosidase (EC 3.2.1.20) - Aspergillus oryzae
C:Species: Aspergillus oryzae
C:Date: 21-Nov-1995 #sequence revision 08-Feb-1996 #text_change 21-Jul-2000
C:Accession: JC4217
R:Minetoki, T.; Gomi, K.; Kitamoto, K.; Kumagai, C.; Tamura, G.
Biosci. Biotechnol. Biochem. 59, 1516-1521, 1995
A:Title: Nucleotide sequence and expression of alpha-glucosidase-encoding gene (agda)
A:Reference number: JC4217; MUID:96032211; PMID:7549103
A:Accession: JC4217
A:Molecule type: DNA
A:Residues: 1-985 <MIN>

```



Qy	812	LVRPAESGIIINLITSSVRNGVAMSQRW----	VMDKITTILGLKRR----	VKIKEYTVQK	865
Db	897	----YPETLSVDPLASRSTLRASARG--TWKEANPLANVLGVTEKPSVTINGETLSS	951		
Qy	865	D-----	AGAIKVKGLGRRTS	879	
Db	952	DSVKYNATSHVLHVGGLOKHITA	973		

RESULT 14

UUHU

sucrose alpha-glucosidase (EC 3.2.1.48) / oligo-1,6-glucosidase (EC 3.2.1.10) [valida  
N:Alternate names: limit dextrinase; small intestinal sucrase/isomaltase (SI)  
C:Species: Homo sapiens (man)  
C:date: 19-Nov-1988 #sequence\_revision 24-May-1996 #text\_change 08-Dec-2000  
C:Accession: S36082; A27326; S24329; A61136  
R:Lacasa, M.  
submitted to the EMBL Data Library, December 1991  
A:Reference number: S36082  
A:Accession: S36082  
A:Molecule type: mRNA  
A:Residues: 1-1827 <LAC>  
A:Cross-references: EMBL:X63597; NID:g36644; PIDN:CAA45140.1; PID:g36645  
R:Green, F.; Edwards, Y.; Hauri, H.P.; Povey, S.; Ho, M.W.; Pinto, M.; Swallow, D.  
Gene 57, 101-110, 1987  
A:title: Isolation of a cDNA probe for a human jejunal brush-border hydrolase, sucra  
A:Reference number: A27326; MUID:88112852; PMID:2962903  
A:Accession: A27326  
A:Molecule type: mRNA  
A:Residues: 1-661, 'X', 663-678 <GRE>  
A:Cross-references: GB:M22616  
R:Chantrel, I.; Lacasa, M.; Chevalier, G.; Ruf, J.; Islam, I.; Mantei, N.; Edwards, Y.  
Biochem. J. 285, 915-923, 1992  
A:title: Sequence of the complete cDNA and the 5' structure of the human sucrase-isom  
A:Reference number: S24329; MUID:92359963; PMID:1353958  
A:Accession: S24329  
A:Molecule type: mRNA  
A:Residues: 1-661, 'F', 663-931 <CHA>  
A:Cross-references: EMBL:X63597  
R:Gorvel, J.P.; Ferrero, A.; Chambraud, L.; Rigal, A.; Bonicel, J.; Maroux, S.  
Gastroenterology 101, 618-625, 1991  
A:title: Expression of sucrose-isomaltase and dipeptidylpeptidase IV in human small i  
A:Reference number: A61136; MUID:91317403; PMID:1677636  
A:Accession: A61136  
A:Molecule type: protein  
A:Residues: 2-14, 'F', 16-20; 1008-1015, 'E', 1017-1021, 'TX', 1024 <GOR>  
C:Genetics:  
A:Gene: GDB:SI  
A:Cross-references: GDB:120377; OMIM:222900  
A:Map position: 3q25.2-3q26.2  
C:Complex: the two product chains remain associated after cleavage  
C:Function: <ISM>  
A:Description: oligo-1,6-glucosidase catalyzes the hydrolysis of 1,6-alpha-D-glucosid  
A:Pathway: carbohydrate digestion  
C:Function: <SUC>  
A:Description: sucrose alpha-glucosidase catalyzes the hydrolysis of the 1,2-alpha-D-  
A:Pathway: carbohydrate digestion  
C:Superfamily: sucrase/isomaltase  
C:Keywords: carbohydrate digestion; duplication; glycoprotein; glycosidase; hydrolase  
F:2-1007/Product: oligo-1,6-glucosidase (isomaltase chain) #status experimental <MARI  
F:13-12/Domain: membrane associated #status predicted <TMW>  
F:42-60/Region: serine/threonine-rich  
F:63-109/Domain: trefoil homology <TRF1>  
F:189-840/Domain: sucrase/isomaltase homology <SIM>  
F:931-977/Domain: trefoil homology <TRF2>  
F:1008-1827/Product: sucrose alpha-glucosidase (sucrase chain) #status experimental <  
F:1062-1734/Domain: sucrase/isomaltase homology <SIM2>  
F:12/Binding site: carbohydrate (Ser) (covalent) #status absent  
F:99-437,458,823,855,904,926,1235,1303,1340,1354,1403,1535,1572,1675,1748,1763,1815/AT  
F:1007-1008/Cleavage site: Arg-Ile (trypsin) #status predicted  
F:1007-1008/Cleavage site: Arg-Ile (trypsin) #status predicted



Query Match	26.8%	Score 1304	DB 1	Length 1827
Best Local Similarity	34.6%	Pred. No. 1.7e-83		
Matches 311	Conservative 157	Mismatches 316	Indels 114	Gaps 28
QY	43	YGVOYNAKVDNSGKSLTALLQILRNSPVYGPDIHFLSFSTASEEDDTLRIRETDDNNR	102	
DB	112	HGVNQDM---TTTISIGVEAKLNIRPSTLFGNDINSLVFTQNTQNRFRFKITDNNR	168	
QY	103	RWEIPNEVLPRPPPPPPPLSSLIQHLPKPIQKQOPTTVLSHSHSDLAFTLPHFTIPGF	162	
DB	169	RYEVPQYVKEFTGFTVSDTLYDVK-----VAQN-----PFSI	201	
QY	163	TIYKSTHDLVLDATPIPSNPTTELIYKQYQLQLSSLPAAQAAHLYGLGEHTKTFGLAH	222	
DB	202	QVIRKSNKGLTFDTSIGF-----LVYSQYQLQISARLPSD--YIYGIGEOVHKRFR--H	251	
QY	223	NQILTLWNAIASPNRL-----NLYGSHPPFMDVRSSPMVGSTHGVFLLNSGHDVE	275	
DB	252	DLGSKWT---PIFTRDQLPGDNNNLVGHQTFMCIEDTS--GKSFQGVFLNNSNAWEIF	305	
QY	276	YTGDRI--TYKVIIGIIDLTYFAGRTPEMLDQYTKLIGRPAPMPYMAPGFHQCRGWYRDV	334	
DB	306	IQTPVITYRVTGILDFYLLGDTPEQVVOYQOLVGLPAMPAYMNLGFLSRWNTKSL	365	
QY	335	NEIETVYDKYAEARIPLEVMMWTDIDYMDAFKDFTLDPVHFPLDKMOQFVTKLHNGORYV	394	
DB	366	DVVKVVRNRREAGIPEDQTQVDIDYEDKKOFTYQVAF--NGLPQFQDLHDHGQKYV	423	
QY	395	PILDPGINTWK-----SYGPIRQMOSNVFIKRN--GNPYLGSVWPGVYVPDFLDPAA	447	
DB	424	IILDPAISIGRRANGTYYATVERGNTQHVWMINESDGGSTPIIGEVMPGLTVVPDTNPNC	483	
QY	448	SFWDEIKRFRDILPIDGIMTDMKEANSFITSAPTG---STLDNPPYKINNNSGRVYP--	502	
DB	484	DMANEGSIFHOEVQYQDGLWIDMNEVSSFQGS--TKGCNVKNLNYPP-----TPDI	534	
QY	503	----INSKTIPTAM--HYGNVTEYNANLNYGLESQATREAL--VRPATRGPFLLSRSTF	555	
DB	535	LDXLMYSKTIICMDAVQNWKG--QYDVHSLYGYSMAIATEQAVQKVFNNKRS--FILTRSTF	591	
QY	556	AGSKYTAHWGTGONARWDDLOYSIPTMLNFGLFMPMIGADICGFABSTTEELCCRWIQ	615	
DB	592	AGSGRAAAHLWGNTASWEQMEWSITGMLEPFLSGIPLVGADICGFVAETTEELCCRWMQ	651	
QY	616	LGAFYPSRSHSARDTHQELYLW---ESVAASARTVLGLRYELLPPYYTYLWYDANLRGS	672	
DB	652	LGAFYPSRHNNSDGYEQHDPAFFGQNSLLVKSRQVLYTHYLLPELYTLFYKAHFGE	711	
QY	673	PIARPLFTFPDDVATYGISQFLLIGRIMVSPVLQPGSSIVNAYSPRGNWVLSNYSSTSS	732	
DB	712	TVARPVLHFEVDNTSNIEDTEFLWGPALLTIPVLKQADTVSYAIPDAIY---DYESG	768	
QY	733	VSVS--AGTYVYSLSAPPHINVHIHEGNIVAMQGEAMTTQAAARSTPFHLLVWMSDHVASTG	791	
DB	769	AKRPMRQRVDMYLPADKIGHLRGGYIIPQEPDVTITASRKNPLGLIIVALGENNTAKG	828	
QY	792	ELFLDNGIEMDIGGPGKWTLVRRFFABSGINNLTISSEVNRGVAMSORWYMDKITLGL	851	
DB	829	DFWDDGETKDT-IQNGWYILYTF-----SVSNNTLIDVCTHSSYQEGTTFLAFQTVKILGL	883	
QY	852	KRRVKIKEYTVQKDAGAIKVKGLGRRTSSHNQGGFFVS-----VTSDLRQLVQQAFLKE	905	
DB	884	TDSVT-----EYRVAAENQPMNAHNSFTYDASNQVLLIADUKLNLGRNFSVQ	930	

RESULT 15

T10799

sucrose alpha-glucosidase (EC 3.2.1.48) / oligo-1,6-glucosidase (EC 3.2.1.10) - rat

C;Species: Rattus norvegicus (Norway rat)

C:\Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000

C;Accession: T10799

R;Chandrasena, G.; Osterholm, D.E.; Sunitha, I.; Henning, S.J.

Gene 150, 355-360, 1994

A>Title: Cloning and sequencing of a full-length rat sucrose-isomaltase-encoding cDNA

A:Reference number: Z17155; MUID:95121929; PMID:7821806

A:Accession: T10799

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1841 <CHA>

A:Cross-references: EMBL:L25926; NID:G414818; PION:AAA65097.1; PID:g773669

A:Experimental source: strain Sprague-Dawley, intestine

C:Genetics:

A:Gene: SI

C:Function:

A:Description: sucrose alpha-glucosidase catalyzes the hydrolysis of the 1,2-alpha-D-:

A:Pathway: carbohydrate digestion

C:Superfamily: sucrose/isomaltase; sucrose/isomaltase homology; trefoil homology

C:Keywords: carbohydrate digestion; glycosidase; hydrolase

F:139-846/Domain: sucrose/isomaltase homology <SIM>

F:937-983/Domain: trefoil homology <TRF>

Query Match	26.6%	Score 1292.5	DB 2	Length 1841
Best Local Similarity	35.7%	Pred. No. 1.1e-82		
Matches	304	Conservative 142	Mismatches 295	Indels 111
Gaps	29			
Qy	40	AICGYGVNNAKVDNSTGKSLTALLQIRNSPVYGPDIHFLSTFASFEEDDTLIRIF	TD	99
Db	119	ADNRHGNAESITNENA--GLKATLNRIPTLFGEDIKSVILTTQTQGNRFRFKITDP		175
Qy	100	NNRRWBIPEVLPRPPPPSPPLSSLOLHPKPIQNPQNTTIVLSHPHSDLAFTLPH	HTP	159
Db	176	NNKREYVPHQVKEETGIPAADLTLYDQ-----VSEN-----P		208
Qy	160	FGTIVRKSTHVDLVPATPIPSNPTTFLYKXOYLQLSLSPAAQAHLYGLGHTKPT	Q	219
Db	209	FSIKVIRKSNKVKLCDSVGP-----LLYSNOYLQISTRLPSE--YIYGF	GHIH	260
Qy	220	LAHQIILTLNADIASFNRD-----LNLGSHPEYMDVRSSPMVGSYTHGVPL	LSNGM	272
Db	261	--HDLVYKWTW-----PIFTKDEIPGDNHNLHGHTFFMGIGDTS--GKSYGV	FELMNSNAM	312
Qy	273	DVEYTGDR-ITYKVIIGIIDLXIFAGRTPEMVLDDQYTKLIGRAPMYPWAFG	PHQC	331
Db	313	EVFIQTPITIVYTGGLDFYFLGDTPEQVVQOYQEVHWRPAMPAYWNLG	FQLSRWNY	372
Qy	332	RDVNEIETVVDKYAERIPLEVMVMTDIDYMDAFKFTLDPVHFPLDKMOQFV	TKLHRNQ	391
Db	373	GLSDTVSEVVRNRREAGIPYDAQVTDIDYMDHKEFTYDRVKF--NGLPE	FAQDLHNHG-	429
Qy	392	RYVPILDPGINTNK-----SYGTFIRGMSNVPIKENG--NPVLGSVWPCPVY	PDPLDP	444
Db	430	KYIILDPALISINKRANGAEYQYTVRGNKQVWVNESDGTTPLIGEVWGLTV	YPDFTNP	489
Qy	445	AARSFWDEIKRFRDILPIDGIMIDWNEASNFITSA-----P-TPG	STLDN	489
Db	490	QTIEWANECNLPHQOQVEYDGLMIDWNEVSSFLQGSINLKGVLIVLVNPP	PTFG-ILDK	548
Qy	490	PPYKINNSGRVPINSKITPATAM-HYGNVTEYNAHNLYGFLESQATREALVR--PATRG		546
Db	549	VMY-----SKTLCMDAIVOHGK-QYDVHSLYGYSMIAITEQAVERVF	FNKRS	594
Qy	547	PFLLSRSTFAGSKYTAHWTDGNAARDLQVSIPTMLNFGIFGMPMICADICG	FAESTT	606
Db	595	-FLTRSTFGGSRHANHMLGDNATASQEWMSITGMLEFGIFGMPVLGATSG	CGFLADTT	653
Qy	607	EEUCCRWIOLGARYPFSRDHSARDTHQE-LYLWESVAASARTVLGLRYELL	PPYVTLMY	665
Db	654	EEUCRMMQLGARYPTSRNNHNAEGMEQDPAYPGQD---SSRHYLTIRTL	PLFLTLFY	710
Qy	666	DANLRGSPARPLSTFPDDVATYGISOFFLIGRGIMVSPVLQGSSIVNAYS	PRGNWYS	725
Db	711	RAHMFGETVARPLYEYFDYDNTSWIEDTQFLWGPALLIITPVLPRGVEN	YSAYIPNATWY-	769
Qy	726	LSNYTSSVSYS-AGTYVYSIAPPDHHINVHIHEGNIIVAMQGEAMTTOA	ARSTPFLHLVWMS	784

Db 770 --DYTGIRPWRKERINMYLPQDKIGLHLRGGYIIPTOEPDVTTTASRKNPLGLIVALD 827  
Qy 785 DHVASTGELFDNGIEMDIGOPGCKWTLVRFFFAESGINNLTISSEVVRGYANSQRWMD 844  
Db 828 DNQAAKGELFWDDGESKD--SIEKMYILYTFVSVN-NELVLN--CTHSSYAEGTS:AFK 882  
Qy 845 KITILGLKRRVK 856  
Db 883 TIKVLGLREDVR 894

Search completed: October 27, 2003, 10:22:06  
Job time : 28.8921 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:20:43 ; Search time 109.802 Seconds  
(without alignments)  
1393.955 Million cell updates/sec

Title: US-10-043-418-2

Perfect score: 4862

Sequence: 1 MERSKLPRICTLAVALPL.....RLVGQAFKLELEFRGATRV 914

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 629382 seqs, 167460630 residues

Total number of hits satisfying chosen parameters: 629382

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4862	100.0	914	14	US-10-043-418-2
2	3243.5	66.7	903	14	US-10-043-418-3
3	2719.5	55.9	902	14	US-10-043-418-4
4	2370	48.7	877	14	US-10-043-418-1
5	1895	39.0	707	12	US-10-032-189-125
6	1304	26.8	1827	9	US-09-819-247-2
7	853.5	17.6	914	12	US-10-032-189-42
8	850.5	17.5	912	12	US-10-032-189-44
9	849.5	17.0	914	12	US-10-032-189-40
10	828	17.0	966	12	US-10-032-189-121
11	823	16.9	943	12	US-10-032-189-123
12	823	16.9	944	12	US-10-032-189-122
13	821	16.9	966	12	US-10-032-189-120
14	818	16.8	944	12	US-10-032-189-134
15	802.5	16.5	967	12	US-10-032-189-38

16	744	15.3	565	15	US-10-102-806-557	Sequence 557, Appl
17	696	14.3	693	12	US-10-228-063-5	Sequence 5, Appli
18	696	14.3	712	12	US-10-228-063-27	Sequence 27, Appl
19	696	14.3	718	12	US-10-228-063-26	Sequence 26, Appl
20	696	14.3	718	12	US-10-228-063-36	Sequence 36, Appl
21	600.5	12.4	788	15	US-10-156-761-14497	Sequence 14497, A
22	576.5	11.9	235	9	US-09-734-569-132	Sequence 132, App
23	509	10.5	1070	10	US-09-280-197-6	Sequence 6, Appli
24	509	10.5	1070	10	US-09-423-126-4	Sequence 4, Appli
25	498.5	10.3	1066	10	US-09-280-197-5	Sequence 5, Appli
26	498.5	10.3	1066	10	US-09-423-126-3	Sequence 3, Appli
27	415	8.5	642	15	US-10-156-761-10107	Sequence 10107, A
28	359	7.4	188	9	US-09-734-569-134	Sequence 134, App
29	347.5	7.1	1092	10	US-09-423-126-5	Sequence 5, Appli
30	336	6.9	570	10	US-09-423-126-6	Sequence 6, Appli
31	317	6.5	1088	10	US-09-280-197-1	Sequence 1, Appli
32	317	6.5	1088	10	US-09-423-126-1	Sequence 1, Appli
33	309	6.4	1091	10	US-09-280-197-2	Sequence 2, Appli
34	309	6.4	1091	10	US-09-423-126-2	Sequence 2, Appli
35	301	6.2	199	9	US-09-734-569-130	Sequence 130, App
36	124	2.6	317	10	US-09-280-197-9	Sequence 9, Appli
37	119	2.4	740	12	US-10-032-585-7673	Sequence 7673, Ap
38	118.5	2.4	785	11	US-09-733-643-2	Sequence 2, Appli
39	117	2.4	682	9	US-09-798-791-2	Sequence 2, Appli
40	114	2.3	1569	12	US-10-287-274-312	Sequence 312, App
41	113.5	2.3	2301	11	US-09-822-871-4	Sequence 4, Appli
42	112.5	2.3	980	12	US-10-320-351-9	Sequence 9, Appli
43	109.5	2.3	630	15	US-10-081-872-202	Sequence 202, App
44	106.5	2.2	1044	10	US-09-960-226-3	Sequence 3, Appli
45	106.5	2.2	2383	9	US-09-912-020-302	Sequence 302, App

ALIGNMENTS

RESULT 1  
US-10-043-418-2  
; Sequence 2, Application US/10043418  
; Publication No. US2002018462A1  
; GENERAL INFORMATION:  
; APPLICANT: Henson, Cynthia A.  
; APPLICANT: Muslin, Elizabeth H.  
; TITLE OF INVENTION: Modified barley alpha-glucosidase  
; FILE REFERENCE: 960296.97486  
; CURRENT APPLICATION NUMBER: US/10/043.418  
; CURRENT FILING DATE: 2002-06-25  
; PRIOR APPLICATION NUMBER: 60/260,787  
; PRIOR FILING DATE: 2001-01-10  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 914  
; TYPE: PRT  
; ORGANISM: Sugar beet  
US-10-043-418-2

Query Match		100.0%;	Score 4862;	DB 14;	Length 914;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 914;		Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
Qy	1	MERSKLPRICTLAVALPLVLCMVVEGATTSKNDQGEAIGYGVQVNAKVDNSTGKSL	60		
Db	1	MERSKLPRICTLAVALPLVLCMVVEGATTSKNDQGEAIGYGVQVNAKVDNSTGKSL	60		
Qy	61	TALLQLTNSPVYCPDTHFLSFTASFEEDDTLRIRFTDANRRWEIPNEVLPRPPPPSP	120		
Db	61	TALLQLTNSPVYCPDTHFLSFTASFEEDDTLRIRFTDANRRWEIPNEVLPRPPPPSP	120		
Qy	121	PPLSSLOHLPKPONOPTTTLVLSHPSHDLAFTLFTHTTPFGFTYRKSTHDLVFDATPIP	180		
Db	121	PPLSSLOHLPKPONOPTTTLVLSHPSHDLAFTLFTHTTPFGFTYRKSTHDLVFDATPIP	180		

```
181 SNPTFTLIYKDOYLQQLSSLPAAQAHLYGLGEHTKPTFQLAHNOILTLWNADIASFNRLD 240
181 SNPTFTLIYKDOYLQQLSSLPAAQAHLYGLGEHTKPTFQLAHNOILTLWNADIASFNRLD 240
241 NLXGSHPFYMDVRSSPMVSGTHGVFLNSGMDVEYTGDRITYKVIIGIIDLIFAGRTIP 300
241 NLXGSHPFYMDVRSSPMVSGTHGVFLNSGMDVEYTGDRITYKVIIGIIDLIFAGRTIP 300
301 EMVLDOYTKLIGRAPMPYAFGHQCRWGYRDVNEIETVVDKYAEARIPLEVMMTDIDY 360
301 EMVLDOYTKLIGRAPMPYAFGHQCRWGYRDVNEIETVVDKYAEARIPLEVMMTDIDY 360
361 MDAFKDFTLPDVHPPLDKMOQFVTKLHRNGORYVPIIDPGINTNKSYGTPIRQMQSNVF 420
361 MDAFKDFTLPDVHPPLDKMOQFVTKLHRNGORYVPIIDPGINTNKSYGTPIRQMQSNVF 420
421 KRNGNPYLGSMVPGVYVYDFLDPAAASFVWDEIKRFRDILPIDGIWIDMNEASNFITSA 480
421 KRNGNPYLGSMVPGVYVYDFLDPAAASFVWDEIKRFRDILPIDGIWIDMNEASNFITSA 480
481 PTGSLTDNPPYKINNSGGVPIINSKTIPTAMHYGNVTEYNNAHNLGYFLESQATREALV 540
481 PTGSLTDNPPYKINNSGGVPIINSKTIPTAMHYGNVTEYNNAHNLGYFLESQATREALV 540
541 RPATRGPFLLSRSTFAGSGKYTAHTWTDGNAARWDDLOYSIPTMLNFGLFQMPMIGADICG 600
541 RPATRGPFLLSRSTFAGSGKYTAHTWTDGNAARWDDLOYSIPTMLNFGLFQMPMIGADICG 600
601 FAESTTELCRRWILQAGAFYFSDHSDHARDTHQELYLWESVAASARTVLGLRYELLPPY 660
601 FAESTTELCRRWILQAGAFYFSDHSDHARDTHQELYLWESVAASARTVLGLRYELLPPY 660
661 YTLMYDANLRGSPPIARPLSFTEPDVATYGISOFLIGRGMVSPVLQPGSSIVNAYSRR 720
661 YTLMYDANLRGSPPIARPLSFTEPDVATYGISOFLIGRGMVSPVLQPGSSIVNAYSRR 720
721 GNWVSLSNYTSVSSVSAGTYVSLSAPPDHNHVIHEGNIIVAMQGEAMTQAARSTPFHLL 780
721 GNWVSLSNYTSVSSVSAGTYVSLSAPPDHNHVIHEGNIIVAMQGEAMTQAARSTPFHLL 780
781 VMSDHDVASTGELFLDNGIEMDIQKAGAIKVKGLGRRTSSHNOGPFVSVISDLROLVQ 840
781 VMSDHDVASTGELFLDNGIEMDIQKAGAIKVKGLGRRTSSHNOGPFVSVISDLROLVQ 840
841 WVMKDITILGLKRRVKIKETVQKDGAIKVKGLGRRTSSHNOGPFVSVISDLROLVQ 900
841 WVMKDITILGLKRRVKIKETVQKDGAIKVKGLGRRTSSHNOGPFVSVISDLROLVQ 900
901 AFKLELEFEGATRV 914
901 AFKLELEFEGATRV 914
```

## RESULT 2

```
US-10-043-418-3
; Sequence 3, Application US/10043418
; Publication No. US20020184662A1
; GENERAL INFORMATION:
; APPLICANT: Henson, Cynthia A.
; APPLICANT: Muslin, Elizabeth H.
; APPLICANT: Clark, Suzanne E.
; FILE OF INVENTION: Modified barley alpha-glucosidase
; TITLE REFERENCE: 960296.97486
; CURRENT APPLICATION NUMBER: US/10/043,418
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: 60/260,787
; PRIOR FILING DATE: 2001-01-10
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 903
; TYPE: PRT
; ORGANISM: Spinach
```

US-10-043-418-3

```
Query Match      66.7%; Score 3243.5; DB 14; Length 903;
Best Local Similarity 67.3%; Pred. No. 7.5e-275;
Matches 610; Conservative 122; Mismatches 166; Indels 9; Gaps 5;
```

```
QY 3 RSKLPRYICPTLAVVLPLVLCMVVEGATTSKNDNGEAGIYGIVYQVAKVDNSTGKSLTA 62
DB 2 KKKIPSL---ALGILVFLVLLQYLQVLVAGISTENSEDEG--VIGYGVKVKVSDSGTFRSLTA 57
QY 63 LLOQIRNSPVYGPDIHFLSFSTASFEEDDTLIRIRTDANRRRWEIPNEVLPR-PPPPSP 121
DB 58 LPQLVKNSSVYGPDIQLLSITASLESNDRLVRITDAKRRWEIPDNILHRHQPPPPPH 117
QY 122 PLSSLOHLXPDIPONQPTTVLSSHPSOLAPTLPHITPFGFTIYRKSTHVDLPDATPIS 181
DB 118 SLSSLYRTLSSPTNRRKILLSPNSDLTFLSLINTTTPPGFTISRKSTHVDLPDATPDT 177
QY 182 NPTTFLIYKDOYLQQLSSLPAAQAHLYGLGEHTKPTFQLAHNOILTLWNADIASFNRLD 241
DB 178 NPNTFLIFIDYLLHLSLPGTRAHIYGLGEHSKPTFQLAHNOILTLWRAADIPSSNPDVN 237
QY 242 LYGSHPFYMDVRSSPMVSGTHGVFLNSGMDVEYTGDRITYKVIIGIIDLIFAGRTPE 301
DB 238 LYGSHPFYMDVRSSPMVSGTHGVFLNSGMDVEYTGDRITYKVIIGIIDLIFAGRTPE 297
QY 302 MYLDOYTKLIGRAPMPYAFGHQCRWGYRDVNEIETVVDKYAEARIPLEVMMTDIDYM 361
DB 298 QVBEQFTVIGRAPMPYAFGHQCRWGYRDVNEIETVVDKYAEARIPLEVMMTDIDYM 357
QY 362 DAFKDFTLDPVHPPLDKMOQFVTKLHRNGORYVPIIDPGINTNKSYGTPIRQMQSNVF 421
DB 358 DAYKDFTLDPVHPPLDKMOQFVTKLHRNGORYVPIIDPGINTNKSYGTPIRQMQSNVF 417
QY 422 RGNPNYLGSMVPGVYVYDFLDPAAASFVWDEIKRFRDILPIDGIWIDMNEASNFITSA 481
DB 418 RGNPNYLGSMVPGVYVYDFLDPAAASFVWDEIKRFRDILPIDGIWIDMNEASNFITSA 477
QY 482 TPGSLTDNPPYKINNSGGVPIINSKTIPTAMHYGNVTEYNNAHNLGYFLESQATREALV 541
DB 478 IPGSTLDNPPYKINNSGGVPIINSKTIPTAMHYGNVTEYNNAHNLGYFLESQATREALV 537
QY 542 PATRGPFLLSRSTFAGSGKYTAHTWTDGNAARWDDLOYSIPTMLNFGLFQMPMIGADICG 601
DB 538 LTERKPFVLSRSTFAGSGKYTAHTWTDGNAARWDDLOYSIPTMLNFGLFQMPMIGADICG 597
QY 602 AESTTELCRRWILQAGAFYFSDHSDHARDTHQELYLWESVAASARTVLGLRYELLPPY 661
DB 598 LGNTTEELCREWILQAGAFYFSDHSDHARDTHQELYLWESVAASARTVLGLRYELLPPY 657
QY 662 TLMYDANLRGSPPIARPLSFTEPDVATYGISOFLIGRGMVSPVLQPGSSIVNAYSRR 721
DB 658 TLMYDANLRGSPPIARPLSFTEPDVATYGISOFLIGRGMVSPVLQPGSSIVNAYSRR 717
QY 722 GNWVSLSNYTSVSSVSAGTYVSLSAPPDHNHVIHEGNIIVAMQGEAMTQAARSTPFHLL 781
DB 718 NWFDLFDYTRSVTASTGRYVTLSPDHNHVIHEGNIIVAMQGEAMTQAARSTPFHLL 777
QY 782 VMSDHDVASTGELFLDNGIEMDIQKAGAIKVKGLGRRTSSHNOGPFVSVISDLROLVQ 841
DB 778 VMSDHDVASTGELFLDNGIEMDIQKAGAIKVKGLGRRTSSHNOGPFVSVISDLROLVQ 837
QY 842 VMDKITILGLKRRVKIKETVQKDGAIKVKGLGRRTSSHNOGPFVSVISDLROLVQ 899
DB 838 VIDKVTILGLKRRVKIKETVQKDGAIKVKGLGRRTSSHNOGPFVSVISDLROLVQ 895
QY 900 QAPKLEL 906
DB 896 REFKLVL 902
```

RESULT 3

US-10-043-418-4

```
; Sequence 4, Application US/10043418
; Publication No. US20020184662A1
; GENERAL INFORMATION:
; APPLICANT: Henson, Cynthia A.
; APPLICANT: Muslim, Elizabeth H.
; APPLICANT: Clark, Suzanne E.
; TITLE OF INVENTION: Modified barley alpha-glucosidase
; FILE REFERENCE: 960296.97486
; CURRENT APPLICATION NUMBER: US/10/043.418
; PRIOR FILING DATE: 2002-06-25
; PRIOR FILING DATE: 2001-01-10
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 902
; TYPE: PRT
; ORGANISM: Arabidopsis
US-10-043-418-4

Query Match      55.9%; Score 2719.5; DB 14; Length 902;
Best Local Similarity 57.0%; Pred. No. 6.2e-229; Mismatches 219; Indels 25; Gaps 7;
Matches 516; Conservative 146; Mismatches 219; Indels 25; Gaps 7;

Qy 12 PTLAVLPLVLCMVVEGATTSKNDNQGEAIGYGVYQVKNKAKVDNSTGKSLTALLQLIRNSP 71
Db 8 PNIFIV--VVVFSSRRSQVVLBEESTVWGYGVVRSVGV--SNRQVLTAKLDLIKPS 64

Qy 72 VYGPDIHLSFTASFEEDTLIRFTDANNRRHIEPNEVLPRPPPPSPPLSSLOHLPK 131
Db 65 VYAPDIKSLNHLVLSLETSELRIRITDSSQQRWIPETVIPR-----AGNHSR 113

Qy 132 PIPO-----NOPTTTLVLSHPSDLAFTLFTHTPGFTIYRKSTHDLVLDATPIPSPTTF 186
Db 114 RFSTEEDGGSNPNFNFIADSSDLVFLHNTTPFGFSVRSRSGDILFDTPSSSSNTY 173

Qy 187 LIYKQYLQSSSLPAQAAHLYGLGHEKTPFQLAHNQIILTWNADIATSNRDLNLYGSH 246
Db 174 FIFKQFLQSSALPENRSLNLYGIGHEKTSFRLLPGETWTWNADIGSENPDVLYGSH 233

Qy 247 PFYMDVRSS---PMVGSHTGVFLNSGMDVEYTGDRITVYKGIIDLYIFAGRTPEMV 303
Db 234 PFYMDVRSGKNEAGTTHGVLLNSGMDVYEGHRTITVYVGGVLDLYVFAGSPSMV 293

Qy 304 LDQYTKLIGRPAPMPYWAFFGHCRCWGRVDVNETETVVYKYAEARIPILEVWMTDIDYMDA 363
Db 294 MNQYTELIGRPAPMPYWSFGHCRCYGVKVSOLYEVYVGYAKAGIPLEVWMTDIDYMDG 353

Qy 364 FKQFTLDPVHFPDKMQCFVTKLHRNQRYVYVILDPGINTKSYGTGFIHQMSNVFIKRN 423
Db 354 YKQFTLDPVNPEDKMQSFVDTLHKNQKYVLIIDPGIGVDSSYGTYNRGMEADVFIKRN 413

Qy 424 GNPVLSGVWPGVYVYVPOFLDPAARSFWDEIKRFRDILP'DGWIWDMNEASNTSAPT 483
Db 414 GEPYLVGWPGKVFYVPOFLNPAATATFNSIEKMFQEIPLDGLW'DMNEUSNFTSPLSS 473

Qy 484 GSTLDNPPYKINNSGGVPIKNTIPATAMHYGNVTEYNAHNLGYFLESQATREALVRPA 543
Db 474 GSSLDNPPYKINNSGDKPEINNTVPATSIHFNGISEYDAHNLGYLLEAKATHQAVDIT 533

Qy 544 TRGPFLLSRSTFAGSKYTAHTWTDGNAARDLLOYSIPTMLNFGLFQKPMIGADICFAE 603
Db 534 GKRPFLLSRSTFVSSGKYTAHTWTDGNAKMDLAYSIPGILNFGLFQPMVADICGFSH 593

Qy 604 STTELCRWQLGAFYFSDHARQTHQELVYLSVSAASARTVLGLRYELLPPYYTL 663
Db 594 DTTBELCRWQLGAFYFSDHARQTHQELVYLSVSAASARTVLGLRYELLPPYYTL 653

Qy 664 MYDANLRGSPILARPLSFTFPDQVATYGISSQFLGGRGIMVSPVLQPGSSVWYASPRGN 723
Db 654 MYEAVHSGNPIARPLSFTFPDQVATYGISSQFLGGRGIMVSPVLSKQAVADYAFAGNW 713

Qy 724 VLSNYSVSVSAGTYVVSLSAPPDHIINVHIHEGNI VAMQGEAMTTOARSTPHELLVVM 783
Db 724 VLSNYSVSVSAGTYVVSLSAPPDHIINVHIHEGNI VAMQGEAMTTOARSTPHELLVVM 783
```

```
Db 714 PDLFNYSPAVGDSCKHYVRLDTPADHVVNVHVRESIVAMQGEALTTRDARTPYQLLVVA 773
Qy 784 SDHVASTGELFDNGIEMDIQGGPK--WTUVRFAESGINNLTISSEVVNRGVAMSORW 841
Db 774 SRLNISGELFDGGENLRMGAGGGRDWTLVKFRCYVTGKSVVLRSEVVNPEYASKMKW 833
Qy 842 VMDKITILGLKRRVKIKEYTVQKDAGALKVGLGRRTSSHNQGGFFVSV-ISDLRLQVLGQ 900
Db 834 SIGKVTFGFENVENVKTYEVRTSERLRSRISLIKTVDNDPFRFLSVESKLSLLVGK 893
Qy 901 AFKLEL 906
Db 894 KFENEL 899

RESULT 4
US-10-043-418-1
; Sequence 1, Application US/10043418
; Publication No. US20020184662A1
; GENERAL INFORMATION:
; APPLICANT: Henson, Cynthia A.
; APPLICANT: Muslim, Elizabeth H.
; APPLICANT: Clark, Suzanne E.
; TITLE OF INVENTION: Modified barley alpha-glucosidase
; FILE REFERENCE: 960296.97486
; CURRENT APPLICATION NUMBER: US/10/043.418
; PRIOR FILING DATE: 2002-06-25
; PRIOR FILING DATE: 2001-01-10
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 877
; TYPE: PRT
; ORGANISM: Barley
US-10-043-418-1

Query Match      48.7%; Score 2370; DB 14; Length 877;
Best Local Similarity 51.2%; Pred. No. 2.5e-198;
Matches 466; Conservative 147; Mismatches 235; Indels 62; Gaps 15;

Qy 16 VVPLVLCMVVEGATTSKNDNQGEA---IGYGVQVKNK-KVDNST-GKSLTALLQLIRNS 70
Db 7 LLLCLCLCLFAPRLCSKEEGFLAARTVLA VAVTMEGALRAEAATGGRSSTG----- 58

Qy 71 PYGPDHIFLSFTASFEEDTLIRFTDANNRRHIEPNEVLPRP-----PPPSPPP 122
Db 59 -----DVQRLAVYASLETDSRLRVITDADHPREVPODIIIRFAPGVDVLDHAPPASAP 113

Qy 123 LSSQLHLPKPIQNOPTTTLVLSHPSDLAFTLFTHTPGFTIYRKSTHDLVLDATPIPSN 182
Db 114 LOG-----RVLSGASDLVLIV-HASPRFTVSRSTGDTLFDATP----- 153

Qy 183 PTFLLIYKQYLQSSSLPAQAAHLYGLGHEKTPFQLAHNQIILTWNADIATSNRDLN 242
Db 154 ---GLVFRDKYLEVTSALPAGRASLYGLGHEKTSFRLEHNDSTFLWNADIGASVVDNL 210

Qy 243 YGSHPFYMDVRSSPMVGSHTGVFLNSGMDVEYTGDRITVYKGIIDLYIFAGRTPEM 302
Db 211 YGSHPFYMDVRAP---GTAHGVLLLSNGMDVLYGGSVYTYKVGGLDFFFGAGPNPLA 267

Qy 303 VLDOYTKLIGRPAPMPYWAFFGHCRCWGRVDVNETETVVYKYAEARIPILEVWMTDIDYMD 362
Db 268 VDOYTKLIGRPAPMPYWSFGHCRCYGVLSVDSLERVAVYAKARIPLEVWMTDIDYMD 327

Qy 363 AFKQFTLDPVHFPDKMQCFVTKLHRNQRYVYVILDPGINT---NKSYGTGFIHQMSNVF 419
Db 328 GPKQFTLDRVNFATAELRPFVDRLEHNAKQVLIIDPGIRVDPIDATYGTGFRGMOQDIF 387

Qy 420 IYKGNPVLGSGVWPGVYVYVPOFLDPAARSFWDEIKRFRDILPIDGIWDMNEASNFITS 479
Db 388 LKRNGTNFGVNWPGVYVYVPOFLDPAARSFWDEIKRFRDILPIDGIWDMNEASNFYN- 446
```



QY 725 SLSNYTSSVSACTYVYSLAPPDHHVHIEGNIIVAMQGEAMTTOARSTPHLLVMS 784  
Db 579 DL--YTAGEASRGVNTLSAPLKIIPVRVGGSIITQEPALTTESRNPFFLLVALD 636  
QY 785 DHVASTGEFLDNGIEMDGGPGCKWTLVRFFAESGINNLTISSEVNRGYAMSORVMD 844  
Db 637 DNGTASGELYLDDGESIDT--QRGDYLLVQFSAN--NNTLTGTVEVTVGYKXNSTLTLE 691  
QY 845 KITILGL 851  
Db 692 KITILGV 698  
RESULT 6  
US-09-819-247-2  
; Sequence 2, Application US/09819247  
; Patent No. US20010036635A1  
; GENERAL INFORMATION:  
; APPLICANT: Waldman, Scott A.  
; APPLICANT: Park, Jason  
; APPLICANT: Schulz, Stephanie  
; TITLE OF INVENTION: Compositions And Methods For Identifying And Targeting Cancer Cells  
; FILE OF INVENTION: Allimentary Canal Origin  
; CURRENT APPLICATION NUMBER: US/09/819,247  
; PRIOR FILING DATE: 2001-03-27  
; PRIOR APPLICATION NUMBER: 60/192,229  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patent In version 3.0  
; SEQ ID NO 2  
; LENGTH: 1827  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-819-247-2  
Query Match 26.8%; Score 1304; DB 9; Length 1827;  
Best Local Similarity 34.6%; Pred. No. 2.1e-104;  
Matches 311; Conservative 157; Mismatches 316; Indels 114; Gaps 28;  
QY 43 YGYOVKNAKVDNSTGKSLTALLQLIRSPVYGPDIHFLSFSTASFEEDDTLRIRFTDANR 102  
Db 112 HGVNQDM--TTTSGVEAKLNIPSPPTLPGNDINSVLTQNTQNPFRFKITDPNR 166  
QY 103 RWEIPNEVLPRPPPPPPSLSLQLPKPIPNQPTTTLVLSHPSDLATLFLHTPPGF 162  
Db 169 RYEVPHQVKEFTGTVSDTLVDK-----VAQN-----PFSI 201  
QY 163 TYRKSTHDVLFDAITPISNPTTFLYKDQYLQSLSSLPAAQAHLYGLGHEHTKPTFOLAH 222  
Db 202 QVIRKSGKTLFDTSIGP-----LVYSDQYLQISARLPSD--YIYGIGQVHKRFR--H 251  
QY 223 NQILTLNADIASNRDL-----NLYGSHFFVMDVRSSPMVGSTHGVLNNGXDVE 275  
Db 252 DLSWKT-----PIFTRDQLPDNNNNLYGHQTFMCTEDTS--GKSGFVFLMNSAMEIF 305  
QY 276 YTGDRIT--TYKVIIGLIDLYIFAGRTPEWLDQYTKLIGRAPMPYAFGHQCRWGRDV 334  
Db 306 IQPTPIVTVTGILDFYLLGTPTQVOVQYQVLGVPAMPAYMNLGQLSRWNTKSL 365  
QY 335 NEITVVDKYAEARIPLEVMWMTDIDYMDAFKDTLDPVHFPLDRMQOQVTKLRNGQRYV 394  
Db 366 DWKEVVRNRREAGIPPTQVTDIDYMDKDKFTYDQVAF--NGLPQFVQDLHDHGQYV 423  
QY 395 PILDGINTK-----SVGTGIRGQSNVFKRN--GNPYLGSVWPGVYYPDFLDAAR 447  
Db 424 IILDPAISIGRANGTTVATYERNGTQHVWVINESDGSPTIIGEVWPGGLTVYDPFTNPCI 483  
QY 448 SFWDIEIKRFRDILPIDGIMDMNEASFTSPTPG---STLDPVVKINNSGRVY-- 502  
Db 484 DWANECSIFHQVOYDGLWIDMNEVSSFIGS--TKGCNVNKLNYPPF-----TFDI 534

QY 503 ---INSKTIIPATAM--HYGNVTEYNHNLVGLFESQATREAL--VRPATRGPFLLSRSTF 555  
Db 535 LDKMYSKTIOMDAVQNGK--QYDVHSLYGSMAIATEQAQVQKVPNKR--FILTRSTF 591  
QY 556 AGSGYTAHTWTCGDNAARWDDLOYSIPTMLNFGLFQMPMIGADICGFAESTTEELCCRMIQ 615  
Db 592 AGSGRHAARHLGNDTASWEQMEWSITGMLESLFGIPLVGADICGFAETTEELCCRMIQ 651  
QY 616 LGAFYFSDRHSARDTHQELYL---ESVAASARTVLGLRIYELLPPYYITLMDANLGRS 672  
Db 652 LGAFYFSDRHSARDTHQELYL---ESVAASARTVLGLRIYELLPPYYITLMDANLGRS 711  
QY 673 PIARPLSFTFPDDVATYGISSOFLIGRIMVSPVLOPGSSIVNAYSPRGVWVLSNVTSS 732  
Db 712 TVARPVLHEFYEDTNSWIEDTEFLWGPALLITPVLKQADTVSAYIPDAIY---DYESG 768  
QY 733 VSVS--AGTVVSLSAPPDHHVHIEGNIIVAMQGEAMTTOARSTPHLLVMSDHVASTG 791  
Db 769 AKRPWRKQVDMYLPADKIGLHRCGYIPIQEPDVTVTASRKNPLGLI VALGENNTAKG 828  
QY 792 ELFLDNGIEMDGGPGCKWTLVRFFAESGINNLTISSEVNRGYAMSORVMDKITILGL 851  
Db 829 DPFWDGGETKOT--IQNGNYIITYT-----SVSNNTLDIVCTHSSYQEGTTLAPQTVKIILGL 883  
QY 852 KRRVKIKEYTVOKDAGAIKVKGLGRRTSSHQGGFFVS-----VISDLRQLVQGAFAKLE 905  
Db 884 TDSVT-----EYRVAENNOPMNAHSNFTYDASNOVLLIADLKLNLGRNESVQ 930  
RESULT 7  
US-10-032-189-42  
; Sequence 42, Application US/10032189  
; Publication No. US20030170630A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsbrook II, John P  
; APPLICANT: Tchernev, Velizar T  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Zerhusen, Bryan D  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Grosse, William M  
; APPLICANT: Lepley, Denise M  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Grosse, William M  
; APPLICANT: Szekeres, Edward S  
; APPLICANT: Vernet, Corine A.M.  
; APPLICANT: Li, Li  
; APPLICANT: Casman, Stacie J  
; APPLICANT: Boldog, Ferenc L  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gangolli, Esha A  
; APPLICANT: Fernandes, Elma R  
; APPLICANT: Rieger, Daniel K  
; APPLICANT: Edinger, Shlomit R  
; APPLICANT: Gunther, Erik  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Sciore, Paul  
; APPLICANT: Ellerman, Karen  
; APPLICANT: MacDougall, John R  
; APPLICANT: Smithson, Glenda  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-228  
; CURRENT APPLICATION NUMBER: US/10/032,189  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 60/257,495  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/258,171  
; PRIOR FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: 60/269,940  
; PRIOR FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 60/274,192  
; PRIOR FILING DATE: 2001-03-08

```

; PRIOR APPLICATION NUMBER: 60/277,826
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/279,840
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/282,981
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/283,656
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/309,247
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/311,754
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/313,331
; PRIOR FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 260
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 42
; LENGTH: 914
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-032-189-42

```

```

Query Match 17.6%; Score 853.5; DB 12; Length 914;
Best Local Similarity 31.7%; Pred. No. 2.1e-65;
Matches 220; Conservative 117; Mismatches 261; Indels 97; Gaps 24;

Qy 206 HLYGLGHTKPTFOAH---NQILTLNADIASFN--RDNLXGSHPFYMDVRSSPMVGS 260
Db 224 HLYGIQOAE-SHQUKNGGDAVRLNLDVGYQIVDMGIYGVPL---AHLGR 278
Qy 261 THGVFLNSGM-----DVEYT-----GDR--ITYKVGIIIDLYIFAG 297
Db 279 TIGIFLNASLTVETINPEAVEYTLQMGPAVKQKVGSRTHVHMSSEGIIDVLLTG 338
Qy 298 RPEKVLDOYTKLGRPAPMYAFGHCQRMGRYDNNIEITVVDKYAERIPLEVWMTD 357
Db 339 PTPSDVFKQYSHLGTQAMPFLSLGSHQCRWNYEQDKVADGDFEHDIPYDAMWLD 398
Qy 358 IDYMDAPKDFLDPVHRPLDKMQQVTKLHRNGORYVPIIDPGINTKSYGTFLRGQSN 417
Db 399 IEHTGKRYFTWKKRFPNPKRMEL--LRSKKRLVVISDPHIKIDPDYSVVYAKDQG 456
Qy 418 VFIK-RNGNPYLGSGVPGVYYPDLPAPARSPPWDEIKRF-----RDILPIDGIWDM 470
Db 457 FFVKNQEGEDFEGVCMPLSGYLDFTNPKVRE-WYSSLEAFPPVYQGSTDIL--FLWNDM 512
Qy 471 NEANSFITSATPGSTLDNPPYKINNNGRVPINSKTIPTATMHYGNVTEYNAHNLGYFL 530
Db 513 NEPSVF-----RGP--EQTQKNAIHHGNWEHRELHNIYGFY 547
Qy 531 ESQATREALVRPATRG---PFLSRSTPAGSKYTAHTWTDGNAARWDDLOQYSIPTMLNFG 587
Db 548 HQMATAEGLIK-RSKGRPEVLTFRSPAGSKYQAVTGDNTAEWNLKISIPMLLTLS 606
Qy 598 LFGPMIGADICGFAESTTEELCCRWOLGAFYFSDRSHSARDTHOELYLW-ESVAASA 646
Db 607 ITGVSFCADIGGIGNPETELLVRVYQAGAYQFFRCHATMNTKRPEWLFGEETRLI 666
Qy 647 RTVLGLRYELLYVYTLMYDANLRGSPILARPLSFTEPDVATYGISOFLIGRGIYSPV 706
Db 667 REATREYGLLPYWSLYPHAHVASQPVWRPLWFEFFDELTKTFMEDEYLGSSALLHPV 726
Qy 707 LOPGSSIVNAYSPRGN--WVLSNYSNTSVSVSACTYVLSLAPPDHINVHIHEGNVAMQG 764
Db 727 TEPRATTVDVLPGSNEVWY---DYKTAHWEAGCTVKIPVALDTIPVFORGGSVPIK- 782
Qy 765 EAMTQAAARST-----PFHLLVMSDHVASTGELFDLNDGMEIDIGPGGKWLVRFFAE 818
Db 783 ----TTVGKSTGWTMTSESYGLRVALSTKSSVSGELYLDGSHFSQY--LHOKQFLHRKF-- 834
Qy 819 SGINNLTISSEVNRGYSORWYMDKITILGLKR 853
Db 835 SFCSSVLINSFADQRGHYPS-KCVVEKILVLGFRK 868

```

```

RESULT 8
US-10-032-189-44
; Sequence 44, Application US/10032189
; Publication No. US20030170630A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook II, John P
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Patturajan, Meera
; APPLICANT: Grosse, William M
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Shimkets, Richard A
; APPLICANT: Grosse, William M
; APPLICANT: Szekeres, Edward S
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Li, Li
; APPLICANT: Casman, Stacie J
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Gorman, Linda
; APPLICANT: Gangolli, Esha A
; APPLICANT: Fernandes, Elma R
; APPLICANT: Rieger, Daniel K
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Gunther, Erik
; APPLICANT: Millet, Isabelle
; APPLICANT: Sciore, Paul
; APPLICANT: Ellerman, Karen
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glennda
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-228
; CURRENT APPLICATION NUMBER: US/10/032,189
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,495
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/258,171
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/269,940
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 60/274,192
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/277,826
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/279,840
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/282,981
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/283,656
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/309,247
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/311,754
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/313,331
; NUMBER OF SEQ ID NOS: 260
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 44
; LENGTH: 912
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-032-189-44

```

```

Query Match 17.5%; Score 850.5; DB 12; Length 912;
Best Local Similarity 31.5%; Pred. No. 3.8e-65;
Matches 218; Conservative 116; Mismatches 264; Indels 95; Gaps 23;

Qy 206 HLYGLGHTKPTFOAH-NQILTLNADIASFN--RDNLXGSHPFYMDVRSSPMVGS 262

```









```
/ APPLICANT: Vernet, Corine A.M.
/ APPLICANT: Li, Li
/ APPLICANT: Casman, Stacie J
/ APPLICANT: Boldog, Ferenc L
/ APPLICANT: Gorman, Linda
/ APPLICANT: Gangolli, Esha A
/ APPLICANT: Fernandes, Elma R
/ APPLICANT: Rieger, Daniel K
/ APPLICANT: Edinger, Shlomit R
/ APPLICANT: Gunther, Erik
/ APPLICANT: Millet, Isabelle
/ APPLICANT: Sciore, Paul
/ APPLICANT: Ellerman, Karen
/ APPLICANT: MacDougall, John R
/ APPLICANT: Smithson, Glennda
/ TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
/ CURRENT APPLICATION NUMBER: US/10/032,189
/ FILE REFERENCE: 21402-228
/ CURRENT FILING DATE: 2001-12-21
/ PRIOR APPLICATION NUMBER: 60/257,495
/ PRIOR FILING DATE: 2000-12-21
/ PRIOR APPLICATION NUMBER: 60/258,171
/ PRIOR FILING DATE: 2000-12-20
/ PRIOR APPLICATION NUMBER: 60/269,940
/ PRIOR FILING DATE: 2001-02-20
/ PRIOR APPLICATION NUMBER: 60/274,192
/ PRIOR FILING DATE: 2001-03-08
/ PRIOR APPLICATION NUMBER: 60/277,826
/ PRIOR FILING DATE: 2001-03-22
/ PRIOR APPLICATION NUMBER: 60/279,840
/ PRIOR FILING DATE: 2001-03-29
/ PRIOR APPLICATION NUMBER: 60/282,981
/ PRIOR FILING DATE: 2001-04-11
/ PRIOR APPLICATION NUMBER: 60/283,656
/ PRIOR FILING DATE: 2001-04-13
/ PRIOR APPLICATION NUMBER: 60/309,247
/ PRIOR FILING DATE: 2001-07-31
/ PRIOR APPLICATION NUMBER: 60/311,754
/ PRIOR FILING DATE: 2001-08-17
/ PRIOR APPLICATION NUMBER: 60/313,331
/ PRIOR FILING DATE: 2001-08-17
/ NUMBER OF SEQ ID NOS: 260
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 122
/ LENGTH: 944
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/
US-10-032-189-122

Query Match 16.9%; Score 823; DB 12; Length 944;
Best Local Similarity 31.1%; Pred. No. 1e-62;
Matches 221; Conservative 109; Mismatches 279; Indels 102; Gaps 24;

Qy 189 YKDYQLQLASSLPQAQAHYLGHEHTKPTFQAHNQIL-----TLWADTASFN-- 237
Db 239 YGPMVGLDPLFGMB-HVYGIPH-----ADNLRLKVTGEGBPYRLNLDVFQELY 290
Qy 238 RDNLNLYGSHPPFYMDVRSSPMVGSTH---GVFLNLSNGMDVEYTGDRITKYKVG----- 287
Db 291 NPVALYGSVPVL-----LAHNPRLDLGIFWLNAAETWIDISSNTAGKTLFGKMDVQLQ 343
Qy 288 -----GIIDLYIFAGRTPEMVLDQYTKLIGRAPMPYWAFGTHQCRNGY 331
Db 344 GSGETPQTDVRMNSGTGIIIDVFLLLGPSISDVPRQYASLTGTQALPPLFLSLGYHOSRWNY 403
Qy 332 RDVNETTVVDKYAEARIPLEVWMTDIDYMDAFKFTLDPVHVEFLDKMQCFVTKLHRNGQ 391
Db 404 RDAVLEVDQGGDDHNLPCDVIWLDIEHADGGRYFTWDSRFP--QPTMLERLASKRR 461
Qy 392 RYVPILDPGINTKNSXGTGTFIRGMSNVFIK-RNGNPFYLGSMVPGVPVYPPFLDPAARSEW 450
Db 462 KLVAIVDPHIKVDGSRVHVEELNLGLYKTRDGSIDYECWCFGSAGYPDFNTPTMRAW 521

/ 451 VD--EIKFRDILPIDGIWIDMEASNFITSAPTGGSTLDNPPYKINNNGRVPINSKTI 508
/ 522 ANMFSYDNYEGSAPNLFVNDMNEPSVP-----NGP-----EVTM 556
/ 509 PATAMHYGNVTYNAHNLGYFLESQATREALVRPATRG---PFLSRSTFAGSGKYTAHW 565
/ 557 LKDAQHYGGWEHRDVHNIYGLYVHMATADGL-RORSQGMERPFVLAKAFAGSQRFGAVW 615
/ 566 TGDNAARWDDLQYSIPTMLNFGLFQMPMIGADICGFAESTTEELCCRWIOLGAFYFESRD 625
/ 616 TGDNTAEWDHLKISIPMCLSLGLVGLSFCGADVCGFFKNPEPELLVRYMYQMGAYQPPFA 675
/ 626 HSARDTHQELYLWESVAAS-ARTVLGRYELLPYYTYLMDANLGRSPITARLSFTFPD 684
/ 676 HAHLDTGRRPEWLLPSQHNDIIRDALGQRYSLLPFWYTLTYQAHREGIPVYRPLWVQYQP 735
/ 685 DVATYGISSQFLGRGMVSPVLQFGSSIVNAVSP-RGN-WVLSNVTSSVSAGTY-- 740
/ 736 DVTFNIDQYLLGDALLVHPVSDSGAHGVQVYLPQGGEVWYDIQSKYKHGQPT-LYLP 794
/ 741 VLSAPPDHINVHIHEGNIV-AMQGEAMTQAAASTPFHLLVWMSDHSVASTGELFDNGI 799
/ 795 VTLSSIP---VFORGGTIVPRMVRVRSSECMKDDPITLTFVALSPQCTAQGELFLDDGY 850
/ 800 EMDIGPGGKWTLVFFAESGINNLTISSEVVNRYGAMSQRWVMDKITLG 850
/ 851 TPNY--QTRQEFLLRRFSFG--NTLVSSSADPEGHPETPIWI-ERVVIIG 896

RESULT 13
US-10-032-189-120
/ Sequence 120, Application US/10032189
/ Publication No. US20030170630A1
/ GENERAL INFORMATION:
/ APPLICANT: Alsobrook II, John P
/ APPLICANT: Tchernev, Velizar T
/ APPLICANT: Liu, Xiaohong
/ APPLICANT: Spytek, Kimberly A
/ APPLICANT: Zethusen, Bryan D
/ APPLICANT: Patturajan, Meera
/ APPLICANT: Grosse, William M
/ APPLICANT: Lepley, Denise M
/ APPLICANT: Burgess, Catherine E
/ APPLICANT: Shimkets, Richard A
/ APPLICANT: Grosse, William M
/ APPLICANT: Szekeres, Edward S
/ APPLICANT: Vernet, Corine A.M.
/ APPLICANT: Li, Li
/ APPLICANT: Casman, Stacie J
/ APPLICANT: Boldog, Ferenc L
/ APPLICANT: Gorman, Linda
/ APPLICANT: Gangolli, Esha A
/ APPLICANT: Fernandes, Elma R
/ APPLICANT: Rieger, Daniel K
/ APPLICANT: Edinger, Shlomit R
/ APPLICANT: Gunther, Erik
/ APPLICANT: Millet, Isabelle
/ APPLICANT: Sciore, Paul
/ APPLICANT: Ellerman, Karen
/ APPLICANT: MacDougall, John R
/ APPLICANT: Smithson, Glennda
/ TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
/ FILE REFERENCE: 21402-228
/ CURRENT APPLICATION NUMBER: US/10/032,189
/ CURRENT FILING DATE: 2001-12-21
/ PRIOR APPLICATION NUMBER: 60/257,495
/ PRIOR FILING DATE: 2000-12-21
/ PRIOR APPLICATION NUMBER: 60/258,171
/ PRIOR FILING DATE: 2000-12-20
/ PRIOR APPLICATION NUMBER: 60/269,940
/ PRIOR FILING DATE: 2001-02-20
/ PRIOR APPLICATION NUMBER: 60/274,192
/ PRIOR FILING DATE: 2001-03-08
```

1 PRIOR APPLICATION NUMBER: 60/277,826  
2 PRIOR FILING DATE: 2001-03-22  
3 PRIOR APPLICATION NUMBER: 60/279,840  
4 PRIOR FILING DATE: 2001-03-29  
5 PRIOR APPLICATION NUMBER: 60/282,981  
6 PRIOR FILING DATE: 2001-04-11  
7 PRIOR APPLICATION NUMBER: 60/283,656  
8 PRIOR FILING DATE: 2001-04-13  
9 PRIOR APPLICATION NUMBER: 60/309,247  
10 PRIOR FILING DATE: 2001-07-31  
11 PRIOR APPLICATION NUMBER: 60/311,754  
12 PRIOR FILING DATE: 2001-08-17  
13 PRIOR APPLICATION NUMBER: 60/313,331  
14 PRIOR FILING DATE: 2001-08-17  
15 NUMBER OF SEQ ID NOS: 260  
16 SOFTWARE: PatentIn Ver. 2.1  
17 SEQ ID NO 120  
18 LENGTH: 966  
19 TYPE: PRT  
20 ORGANISM: Homo sapiens  
21 US-10-032-189-120

Query Match 16.9%; Score 821; DB 12; Length 966;  
Best Local Similarity 31.1%; Pred. No. 1.6e-62;  
Matches 221; Conservative 109; Mismatches 279; Indels 102; Gaps 24;  
Qy 189 YKQYLOLSSSLPAQAAHLYGLGHTKPTFLAHNOIL-----TLWNADIASFN-- 237  
Db 261 YGPMVGSLDFSLPOME-HVIGIPEH-----ADNLKVKTEGEGPYRLNLDVFOYELY 312  
Qy 238 RDLNLYGSHPFYMDVRSSPMVGSTH---GVFLNSNGMDVEYTGDRITYKVG----- 287  
Db 313 NPWALYGSVPVL-----LAHNPRLDGLFVWLNAAETWVDISSNTAGKTLFGKMMDYIQ 365  
Qy 288 -----GIIDLYFAGTPEWLDQYTKLGRPAPMPYNAFGHQRCWGY 331  
Db 366 GSGETPDVVRMSETGIIDVLLGSPISDVFRQVYASLTGTQALPPLFSLGSHQSRWNY 425  
Qy 332 RDVNEIETVVDKYAEARIPLEVMMVTDIDYMDAFKDFLDVPHFPLDKMQOFVTKLHRNQ 391  
Db 426 RDEADVLEVDGPDHNLPCDVIWLDIEHADGKRYFTWDSRFP--QPRIMLERLASKR 483  
Qy 392 RYVPILDPGINTKSYGTFIRGMSNVFIK-RGNPNYLGWVGPGVYVDFDLPAAARSW 450  
Db 484 KLVAIVDPHIKVDGSGYRVEELRNGLYVTRDGSDEYGCWPGSAGYDPFTNPTRAWW 543  
Qy 451 VD--EIKRERDILPIDGIWIDMNEASFITSAPTPGSTLNDPPYKINNNGRVPINSKTI 508  
Db 544 ANWFSYDNYEGSAPNLFVWMDMNEPSVF-----NGP-----EVTM 578  
Qy 509 PATAMHYGNVTEYNAHNLGFLSEQATREALVRPATRG---PFLSRSTFAGSGKYTAHW 565  
Db 579 LKDAQHYGGWEHRDVHNIYGLYVHMATADGL-RQSGGXEERPFVLARAFAGSQRFCAVM 637  
Qy 566 TGDNAARWDLQVSIPTMLNFGFLGPMIGADTCGFAESTTELCRCRWQLGAFYFPSRD 625  
Db 638 TGDNTAEWDLKHSIPKCLSLGLVGLFCGADYGGFFKNPEPELLVRYWQMGAYQPFRA 697  
Qy 626 HSARDTTHOELYLWESVAAS-ARTVLGRYELLPPYVYTLMYDANLGRGSPARPLSTFPD 684  
Db 698 HAHLDTGRERFWLLPSQHNDIIRDALQRYSLPFPFTLLYQAHREGIPVWRMFWQYFQ 757  
Qy 685 DVATYGISOFFLGRGIMVSPVLPQGSISVNAVSP-RGN-WVLSNYSSTSVSVSAGTY-- 740  
Db 758 DVTTFNIDDDQVLLGDALLVHPVSDSGAHGVQVYLPQGGVEWYDIQSYQKHGHPQT-LYLP 816  
Qy 741 VLSLAPDHNHVIHEGNIV-AMGEMAMTTOARSTPFHLLVWMSORHASTGELFLDNGI 799  
Db 817 VTLSSIP-----VFORGGTIVPRWVRSSSECMDKDDPITLFLVALSPQGTAGGELFLDDGH 872  
Qy 800 EMDIGGFGGKWLVRFFAESGINNLTLSSEVNRVGYAMSORWMDKXITILG 850  
Db 873 TFNY--OTROEFLRRFSFG--NTLVSSADPEGHETIPIWI-ERVVIIG 918

RESULT 14  
US-10-032-189-124  
; Sequence 124, Application US/10032189  
; Publication No. US20030170630A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsobrook II, John P  
; APPLICANT: Tchernev, Velizar T  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Zerhusen, Bryan D  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Grosse, William M  
; APPLICANT: Lepley, Denise M  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Grosse, William M  
; APPLICANT: Szekeres, Edward S  
; APPLICANT: Vernet, Corine A.M.  
; APPLICANT: Li, Li  
; APPLICANT: Casman, Stacie J  
; APPLICANT: Boldog, Ferenc L  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gangolli, Esha A  
; APPLICANT: Fernandes, Elma R  
; APPLICANT: Rieger, Daniel K  
; APPLICANT: Edinger, Shlomit R  
; APPLICANT: Gunther, Erik  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Sciore, Paul  
; APPLICANT: Ellerman, Karen  
; APPLICANT: MacDougall, John R  
; APPLICANT: Smithson, Glennda  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-228  
; CURRENT APPLICATION NUMBER: US/10/032.189  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 60/257,495  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/258,171  
; PRIOR FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: 60/269,940  
; PRIOR FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 60/274,192  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/277,826  
; PRIOR FILING DATE: 2001-03-22  
; PRIOR APPLICATION NUMBER: 60/279,840  
; PRIOR FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/282,981  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: 60/283,656  
; PRIOR FILING DATE: 2001-04-13  
; PRIOR APPLICATION NUMBER: 60/309,247  
; PRIOR FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: 60/311,754  
; PRIOR FILING DATE: 2001-08-17  
; PRIOR APPLICATION NUMBER: 60/313,331  
; NUMBER OF SEQ ID NOS: 260  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 124  
; LENGTH: 944  
; TYPE: PRT  
; ORGANISM: Sus scrofa  
; US-10-032-189-124

Query Match 16.8%; Score 818; DB 12; Length 944;  
Best Local Similarity 31.0%; Pred. No. 2.8e-62;  
Matches 225; Conservative 107; Mismatches 272; Indels 122; Gaps 26;  
Qy 180 PSNPTFLIYKDYQLQLSSSLPAQAAHLYGLGHT-----KPTFLAHNOILTLWNADIA 234





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 ; Search time 16.5211 Seconds  
(without alignments)  
2340.767 Million cell updates/sec

Title: US-10-043-418-2

Perfect score: 4862

Sequence: 1 MERKLPY:PTLAVVLP.....RLVQAFKLEFEGATRV 914

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents\_AA.\*

1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.psp.\*

2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.psp.\*

3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.psp.\*

4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.psp.\*

5: /cgn2\_6/ptodata/2/iaa/PTCUS\_COMB.psp.\*

6: /cgn2\_6/ptodata/2/iaa/backfiles.psp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2370	48.7	877	1	US-08-430-925A-4
2	884.5	18.2	914	4	US-09-437-054A-8
3	816	16.8	919	4	US-09-437-054A-17
4	696	14.3	693	4	US-09-376-343-2
5	681.5	14.0	938	4	US-08-897-843A-1
6	509	10.5	1070	2	US-08-633-770A-2
7	498.5	10.3	1066	2	US-08-633-770A-1
8	347.5	7.1	1092	4	US-09-275-608-3
9	336	6.9	570	4	US-09-275-608-4
10	310	6.4	1091	3	US-08-633-768A-2
11	308	6.3	1088	3	US-08-633-768A-1
12	182.5	3.8	390	4	US-09-634-238-331
13	175.5	3.6	150	4	US-09-437-054A-10
14	119.5	2.5	879	1	US-08-306-546C-2
15	119.5	2.5	879	2	US-08-530-544A-2
16	117	2.4	3892	4	US-09-328-352-5503
17	112.5	2.3	980	1	US-08-091-569-2
18	112.5	2.3	980	1	US-08-203-676-2
19	112.5	2.3	980	2	US-08-822-238-2
20	110	2.3	742	4	US-09-107-532A-4996
21	108	2.2	3060	2	US-08-487-826B-14
22	107.5	2.2	954	3	US-09-251-645-12
23	106	2.2	1596	4	US-09-328-352-5542
24	106	2.2	2285	4	US-09-252-991A-17790
25	105	2.2	566	4	US-09-431-470-2
26	105	2.2	1179	4	US-09-107-532A-6126
27	104.5	2.1	513	3	US-09-041-886-28

28	104.5	2.1	530	3	US-09-041-886-29	Sequence 29, Appl
29	104.5	2.1	552	3	US-09-041-886-30	Sequence 30, Appl
30	104.5	2.1	589	3	US-09-041-886-31	Sequence 31, Appl
31	104.5	2.1	1137	4	US-09-252-991A-24829	Sequence 24829, A
32	104.5	2.1	2628	2	US-08-570-311-14	Sequence 14, Appl
33	104.5	2.1	3144	1	US-08-246-982A-6	Sequence 6, Appl
34	104.5	2.1	3144	1	US-08-453-265-6	Sequence 6, Appl
35	104.5	2.1	3144	2	US-08-457-273B-42	Sequence 42, Appl
36	104.5	2.1	3144	3	US-08-556-419-21	Sequence 21, Appl
37	104.5	2.1	3144	3	US-09-041-886-15	Sequence 15, Appl
38	104.5	2.1	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
39	103.5	2.1	574	3	US-09-383-586-36	Sequence 36, Appl
40	103.5	2.1	633	4	US-09-186-002-18	Sequence 18, Appl
41	103.5	2.1	634	4	US-09-186-002-2	Sequence 2, Appl
42	103	2.1	1289	1	US-07-876-280-4	Sequence 4, Appl
43	103	2.1	1289	1	US-07-675-772-4	Sequence 4, Appl
44	103	2.1	1289	1	US-08-063-170-4	Sequence 4, Appl
45	103	2.1	1289	1	US-08-158-232-4	Sequence 4, Appl

## ALIGNMENTS

RESULT 1  
US-08-430-925A-4  
; Sequence 4, Application US/08430925A  
; Patent No. 5763252  
; GENERAL INFORMATION:  
; APPLICANT: Skadsen, Ronald W  
; APPLICANT: Tibbot, Brian K  
; TITLE OF INVENTION: Cloned Alpha-Glucosidase from Barley  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Quarles & Brady  
; STREET: 1 South Pinckney Street  
; CITY: Madison  
; STATE: WI  
; COUNTRY: US  
; ZIP: 53703  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/430,925A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J  
; REGISTRATION NUMBER: 27,386  
; REFERENCE/DOCKET NUMBER: 960296.93171  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 608-251-5000  
; TELEFAX: 608-251-3166  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 877 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-430-925A-4

Query Match 48.7%; Score 2370; DB 1; Length 877;

Best Local Similarity 51.2%; Pred. No. 4 5e-204;

Matches 466; Conservative 147; Mismatches 235; Indels 62; Gaps 15;

Qy 16 VLVLVLCMVVEGATTSKNDNGEA---IGYGVQVNA-KVDNST-GKSLTALLQLIRNS 70

Db 7 LLILCLCLFAPRLCSSKEEGLAARTVLAIVTMEGALRAEAATGSRSTG----- 58

Qy 71 PVYGPDIHLSTASFEEDDTLRFTDANRRWEIPNEVLPRP-----PPPPSPPP 122



```
Db 59 -----DVQLAVYASLETSRLVRITDADHPRWEVQDIIIPRAPGDVHLHDAPPASSAP 113
Qy 123 LSSLOHLKPIQONQPTTIVLSHPSHDLAFTLHTTTPGFTIYKSTHDLVLFDAITPSPN 182
Db 114 LQG -----RVLSPPAGSDLVLV-HASPERFTVSRRTGDTLFDTPA----- 153
Qy 183 PTFELIYKQVLQSSSLPAQOAHLYLGEGHTKPTFQLAHNQILTLWNADIASFNRLNL 242
Db 154 ---GLVFRDKLEVTSALPAGRASLYLGEGHTKSFRLRHNDSTFLWNADIGASYVDVNL 210
Qy 243 YGSHFFYMDVRSSPMVSGTHGVFLNLSNGMDVEYTGDRITKYVIGGIDLYIFAGRTPEM 302
Db 211 YGSHFFYMDVRAP---GTAHGVLSSNGMDVLYGGSVYTYKVIGGVLDVFFYFAGPNPLA 267
Qy 303 VLDOYTKLIGRAPMPYNAFGHOCRWYRDVNEIETVWKYAEARPLEVMWTDIDYMD 362
Db 268 VDOQTQIARAPAMPYWSFGHOCRYGLNVSDLERVVARYAKARPLEVMWTDIDYMD 327
Qy 363 AFKDFTLDPVHFPKDKMQQFVTKLHRNQRYVPIIDPGINT---NKSYGTIFRGMQSNVF 419
Db 328 GFKDFTLDRVNFATAELRPFVDRHLRHAQKVVILDPGRVDPIDATYGTFRGMQDIF 387
Qy 420 IKRGNPVLGSVMPGPVYVDFLDPAAARSFWDEIKRFRDILPDIGIWDIWEASNFITS 479
Db 388 LKRNGTFVGNVMPGDVYFDFPMHPAAAEFWAREISLFRRTIPVZGLWIDNEISNFYN- 446
Qy 480 APTGSLTLDNPPYKINNSGRVPIINSKTIPTAMHYGNVTEYNAHNLVGFLESQATREAL 539
Db 447 -PEPMNALDDPPYRINNDGTGRPINNKTVRPLAVHYGGVTEYEHNLFGLLERATGRGV 505
Qy 540 VRPATGRPFLLSRSTFAGSGKYTAHWTGDNARMDLQYSIPTMLNFGLFQMPMIGADIC 599
Db 506 LRDTCRRPFVLSRSTFVSGRYTAYMTGDNAAATWCLAYSINTMLS FGLFQMPMIGADIC 565
Qy 600 GFAESTTEELCORWLOLQAFYFSDHDSARDTHOELYLWESVAASARTVLGLRYELLPY 659
Db 566 GFNGNTTEELCORWLOLQAFYFSDHDSARTVRRRELYLWESVAASGRKALGLRYQLLY 625
Qy 660 VYTLMDANLRGSPARPLSFTEPDVATYGISOFLGRGIMVSPVLQSGSIVNAYSP 719
Db 626 FYTLMEAHMTGAPLARPLSFYSDHDAVYGVDRFLGRGVLSVPLDEPGPTTVDAIFP 685
Qy 720 RGNVYSLNYSSTSSVSAGTVLSAPPDHNHVIHEGNIAMQGEAMTTOAARSTPFHL 779
Db 686 AGRWRYLYDSLAVATR-GKHVRLPAPADTVNVHLTGGTILPLOSALTTSRARTAPHL 745
Qy 780 LVVMSDHVASTGELFLDNGIEMDIOGPGKWTLVRFFAESGINN--LTISSEVNVNGYAM 837
Db 746 LVALLAEDGTASGYLFLDDGDSPEY-GRSDMSMVRFNKYIPNNKGAIKVSEVHNRYAQ 804
Qy 838 SORWYMDKITILGLKRRVKIKBYTVQKDAKALKVGLGRRTSSHQGF-FVSVISDLRQ 896
Db 805 SRTLVIKVLGHRSPAPKLTIVHNSAEVASS-SAGTRYQAGGLGVAGHIGGLSL 863
Qy 897 LVGOAFKLEL 906
Db 864 VVGEFPELV 873
```

## RESULT 2

US-09-437-054A-8

; Sequence 8, Application US/09437054A

; Patent No. 6316698

; GENERAL INFORMATION:

; APPLICANT: Allen, Stephen M.

; APPLICANT: Kinney, Anthony J.

; TITLE OF INVENTION: Plant Alpha-Glucosidase II Homologs

; FILE REFERENCE: BB1273 US NA

; CURRENT APPLICATION NUMBER: US/09/437,054A

; PRIORITY FILING DATE: 2001-05-14

; PRIOR APPLICATION NUMBER: 60/107,909

; PRIOR FILING DATE: 1998-No. 6316698member-10

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 8

; LENGTH: 914

; TYPE: PRT

; ORGANISM: Glycine max

; US-09-437-054A-8

Query Match 18.2%; Score 884.5; DB 4; Length 914;

Best Local Similarity 29.7%; Pred. No. 28-70;

Matches 272; Conservative 138; Mismatches 370; Indels 135; Gaps 36;

```
Qy 13 TLAVLPLVLCMVVEGATTSKNDN-----OGEALIGYGYQVKNKAKVDNSTGKS 59
Db 5 TLRLLILLLLSHLHSLVLSWKKEEFRTCHQTPCKARSRAPGSSSLIA-TDVTISHG-D 62
Qy 60 LTRALLQIRNSPVGPDIIHFLSTASPEEDDILIRFTD-----ANRRWEIPNEVLPRP 114
Db 63 LTRALLQIRNSPVGPDIIHFLSTASPEEDDILIRFTD-----ANRRWEIPNEVLPRP 114
Qy 115 PPPSPPLSSLSLQHLPK--PIQONOPTTTLVLSHPSDL---AFTLF-HTTTPGFTIYRK 167
Db 120 P-----STKLMLPKISSVENGLSSSVLSGDHSAVLRHDPFELFIRDSSGDRVISL 171
Qy 168 STHDVLFDAITP-----SNPTFLIYKQVYQLQSSSLPAQOAHLYLGEGHT- 214
Db 172 NSHD-LFDFEQLKHKSEDDNWEQEAFRSHTRRPPYGPQSISFDVSFYGAD-FVYGIPERAA 229
Qy 215 ----KTFP--QLAHNQILTLWNADIASFNRD--LNLVYSHSPFVMDVRSMPVSGTHGVFL 266
Db 230 SLALKTRGPNVDESEPPYRLNLDVEYIHDSPFGLYGSIPFMV---SHGKARGSGGFVW 286
Qy 267 LNSNGMDVE-----YTGDRIT--YKVIIGIIDLIFAGRTPEMVLDQYTKLI 311
Db 287 LNAEQIDVLAPOGWAESSIALPSHRIDTFWMSAAGVVDAPFFIGPNPKDVLRYTAVT 346
Qy 312 GRPAPMPYNAFGHOCRWYRDVNEIETVWKYAEARPLEVMWTDIDYMDAFKDTFLDP 371
Db 347 GTPAMPQLFSLAHQCRWYRDEEDVEHVDKFDLDDIPYDLVLDIEHTDCGRYFTWR 406
Qy 372 VHEP-LDKMQQFVTKLHRNQRYVPIIDPGINTKNSYGTIFRGMQSNVFI-RNGNPFYLG 429
Db 407 ALFPHEEMQR-----KLASKGRHMVTIVDPHKKDENFHLHKEASQKGYVYKDGANDFDG 463
Qy 430 SVMFPGVYVDFLDPAAARSFWDEI--KRFRTDILPDIGIWDIWEASNFITSAPTGSTIL 487
Db 464 WCPFGSSYPDLNPIRSWADKFSYQSVGEGSTPSLYIWNQDNVPSVF----- 512
Qy 488 DNPYPYKINNSGRVPIINSKTIPTAMHYGNVTEYNAHNLVGFLESQATREALVRPATRG- 546
Db 513 -NGP-----EVTMPRDVTHYGVGVEHRELHNAYGYVYFHMATANGLLK---RGE 555
Qy 547 ----PFLSRSTFAGSGKYTAHWTGDNARMDLQYSIPTMLNFGLFQMPMIGADICGFA 602
Db 556 GNDRPFVLSRSTFAGSGKYTAHWTGDNARMDLQYSIPTMLNFGLFQMPMIGADICGFA 602
Qy 603 ESTTEELCORWLOLQAFYFSDHDSARDTHOELYLW-ESVAASARTVLGLRYELLPY 661
Db 616 GNPEPELLVWYQLGAYYPPFRAHAHHTKRRPMLFGERNTELIKDAIHVRVALLPYFY 675
Qy 662 TLMVDANLRGSPARPLSFTEPDVATYGISOFLGRGIMVSPVLQSGSIVNAYSP-R 720
Db 676 TLFREANTTGVVVRPLMPEPDEATFSNDETFFMGSSILVQGIYTERAKHASVVLPGK 735
Qy 721 GNVVLSNYSSTSSVSAGTVLSAPPDHNHVIHEGNIAMQGEAMTTOAARSTPFPH 778
Db 736 QSWYDLR--TGAV-YKGGVTHKLEVTEESIPAFQRTAGTIIARKDRFRSTQMA-NDPYT 791
Qy 779 LLVMSDHVASTGELFLDNGIEMDI--GGPGKWTLVRF--FAESGINNLTTISSEVNVNGY 835
Db 792 LVVALNSSQAEGELYIDGSSNFLOGG---YIHRFFIFSNGKLTSLDAPASSSKGR 847
Qy 836 AMSQRWYMDKITILG 850
| : : : : :|
```



Db 613 ---DELIVYGETSKRYDNAETISSNE---IKFREIYVSKLTITSE 654

## RESULT 5

US-08-897-843A-1  
; Sequence 1, Application US/08897843A  
; Patent No. 6514493  
; GENERAL INFORMATION:  
; APPLICANT: Deleo, Albert B.; Loftus, Douglas; Appella, Ettore  
; TITLE OF INVENTION: CDNA CLONE FOR MURINE TUMOR  
; NUMBER OF INVENTIONS: REJECTION ANTIGEN GP110 AND TUMOR PEPTIDE VACCINE  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Diane R. Meyers  
; STREET: 600 Grant Street, 42nd Floor  
; CITY: Pittsburgh  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 15219  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/897,843A  
; FILING DATE:  
; CLASSIFICATION: 424  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 938 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: NO  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Mus musculus  
; DEVELOPMENTAL STAGE: Embryo  
; TISSUE TYPE: Embryo  
; CELL LINE: NIH 3T3  
US-08-897-843A-1

Query Match 14.08; Score 681.5; DB 4; Length 938;  
Best Local Similarity 27.94; Pred. No. 3.9e-52;  
Matches 207; Conservative 115; Mismatches 284; Indels 157; Gaps 29;

Qy 180 PSNPTFLIYKDYQLQLSSLPAAQAHLYGLGEHT-----KPTFQLAHNQILTLWNADIA 234  
Db 233 PYGPTS-----VGLDFSLPGNE-HVYGIPEHADSRLKVT---EGGEPYRLXNLDVF 280  
Qy 235 SF--NRDLNLYGS-----HPFYMDVRSSPMVSGTHGVFLN-----SN----- 270  
Db 281 QYELNNPMALYGSVPVLLAHSFHRLD-----GIFWLNAAETWVDISSNTAGKTLF 330  
Qy 271 GMOVEY---TGDR---ITVKVIGIIDLIFAGRTPEMVLDXYTKLIGRPAPMPYWAFC 323  
Db 331 GKMDLYLQGGGETPQTDIRMWSGIIIDVFLMLGSPVDFRQVATUTGQALPPLFSLG 390  
Qy 324 FHQRWGYRDVNETETVVDKYAEARILEVWNTIDIDYMDAFKFTLDPVHF--PLDKMQQ 381  
Db 391 YHOSRWNYRDEADVLEVDQGFDDHNMFCVILWIDIEHADGKRYFTWPTTFPQPLANLEH 450  
Qy 382 FVTKLHNGORYVILDPGINTKNSYGTFRIGMQSNVFIK-RGNPNYLVGSWPGVPYYPD 440  
Db 451 LDSK-RNN---VVAIVDPHIKVDSCYRVHBEELNHLGLYVKTRDGSDEYEGWCWPGSAYPD 506  
Qy 441 FLDPAAFSFVVD--EIKFPRDILPDIGIMIDMEASNFITSAPTPGSTLDPNPKYKNNSG 498  
Db 507 FTNPRMALNSMFSFNYEGSAPNLVYMDMNEPSVF-----NGP----- 547

Qy 499 GRVPINSKTIIPATAMHYGNVTEYNAHNLGYFLESOATREALVRPA--TRGPBLLSRSTFA 556  
Db 548 -----EVTMLKDAVHYGGWEHRDINHYYGLYVHMATADGLIQRSGGIERPFVJSRAFFS 601  
Qy 557 GSGKYTAHTWGDNAARMDDLOYSIPTMLNFGLFNPMIGADICGFAESTTEELCCRWIQL 616  
Db 602 GSOREGAVMTGDNTAEWDHLKISIPMCLSLALVGLSGFADVGFFKKNPEPELLVRWYQM 661  
Qy 617 GAFYFPSRDSARDTHOELYLM-----ESVAASA-----RTVLGLRYELLPPYYITLMY 665  
Db 662 GAYGFFFAHAT-----WTLGGESMAVSIVPRNPRCLVPVPAIFFAALLVYPLLS 711  
Qy 666 DANLRGSPARPLSFTFPDDVATYGISQFLLGRIMVSPVLQPGS----- 711  
Db 712 SSQGRVSCHEAPL-VQYFEDMSTFSDQFMGLDALLHPVSDAGHGGRGICIAKRCG 770  
Qy 712 ----SIVNAYSIPRGNWVSLSNYSVSVSACTYVYVLSAPPDHINVHIHEGNIVAMQCEAM 767  
Db 771 MTFRAIRSIMGPRCICPVTLLSSIPVFOGGTIV-----PRMVRV-----R 812  
Qy 768 TQOARSTPFHLLVMSDVHSTGTGELFDNGIBMDIGGPGGKWTIVRFAESGINNLITIS 827  
Db 813 SSDCKMDPITLFLVALSPQGTAGQELFLDDGHTFNY--QTRHEFLRRFSFG--STLVS 868  
Qy 828 SEVNVNRYAMSQRWMDKITILG 850  
Db 869 SSADPKGHLETPIMI-ERVVIMG 890

## RESULT 6

US-08-633-770A-2  
; Sequence 2, Application US/08633770A  
; Patent No. 5308760  
; GENERAL INFORMATION:  
; APPLICANT: Bojlsen, Kirsten  
; APPLICANT: Yu, Shukun  
; APPLICANT: Kragh, Karsten  
; APPLICANT: Christensen, Tove  
; APPLICANT: Marcussen, Jan  
; TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM A FUNGUS, ITS  
; NUMBER OF INVENTIONS: PURIFICATION GENE CLONING AND EXPRESSION IN MICROORGANISMS  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson & Bear  
; STREET: 620 Newport Center Drive 16th Floor  
; CITY: Newport Beach  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/633,770A  
; FILING DATE: July 8, 1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP94/03398  
; FILING DATE: OCT-15-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Altman, Daniel E  
; REGISTRATION NUMBER: 34,115  
; REFERENCE/DOCKET NUMBER: DY0U6.001APC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 714-760-0404  
; TELEFAX: 714-760-9502  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1070 amino acids  
; TYPE: amino acid

```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-633-770A-2

Query Match
Best Local Similarity 10.5%; Score 509; DB 2; Length 1070;
Matches 194; Conservative 108; Mismatches 244; Indels 240; Gaps 34;

QY 159 PFGFTIYKSTHVDLFDATIP-----SNPTTF-----LIYKQDYL 194
DB 153 PFIQVRLT--PLVDPPFIPNVANATARADKVVWQTSPTFRKMLHPQHKLKDTVL 210
QY 195 QLSSSLPAQAQAHLYGLGE-----HTKPTF-----QLAHNQILTLWNADIASFNRLN 241
DB 211 DI---IKFGHGYVGMCGEIEFMKEPTFMVNFDMQYQVY---AQGALDSRE-P 262
QY 242 LYGSHPFYMDVRSSPMVSGTHGVFLNMGDMVEYTGDRITYKVG---GIIDLY-IFA 296
DB 263 LYHSDPPFLVDNSNPEHKNITATFIDNYSQIADFGKTSNGYIKLGRYGGIDCYGISA 322
QY 298 RTPEMVLQDVKTLIGRAPMPYAFGPHQCRGWYRDVNEIETVVDKYAEARIPLEVMWTD 357
DB 323 TVPEIV-RLYTGLVGRSKLPRYLIGHAQACYGQESDLHAVVQQYRDTKFPLDGLHVD 381
QY 358 IDYMDAFKDFLDPVHFPLDKMQQFVTKLHRNGORYVPIIDPGINTN---KSYGTFRGM 414
DB 382 VDFODNFRFTTNPITFPNPK--EMFTNLRNNGIKCSTNITPVISIRDRENGYSTLNERY 439
QY 415 QSNVFIKRN-----GN-----425
DB 440 DKXYFIMDDRYTEGTSQDPPQVRYSFYGGGNPVEVNPNDVMARPDFGNDYDFTNFCKD 499
QY 426 -PYLGSWMPG---PVYYPDFLPPAARSFWDDEIKRFRDILPIDGWIWMNEASNFITS 479
DB 500 YPHGGSYGVGNGTGPYIPDLNREVRIMWGLQYELFN-MGLEFWQDMT-----550
QY 480 APTFGSLDNPYPYKINNSGGRVPINSKTIP-----ATAMHYGNVTE-----YNA 523
DB 551 -TPA-----IHSSYG---DMKGLPTRLVLTADSVTNASEKKLAIESWALYSYNL 595
QY 524 HNL-----YGFLEQATREALVRPATRGPFLLSRSTFAGSGKYTAHWTGDNAAWDDLOYS 579
DB 596 HKATPHGLGRLESRKNR-----NFIILGRSYAGAYRFAGLWTDGNASWEFWFKIS 646
QY 580 IPTMLNGLFGMPMIGADICGFAESTTE-----ELCCRWIQLGAFYFPFSRDSHARD 630
DB 647 VSQVLSLGLNGVCIAGSDTCGFEPARTIEGEEKYCSPELIRWYTSFLLPLWRNHYVXK 706
QY 631 -----TTHQEL-----YLMESVAASARTVLGLRYELLFPYVYTLMYCANLRG 671
DB 707 DRKWFQBPYAYPKHLETHPELADQAMLYKSYLEICRYWVELRYSLIQLLYDCMFQNVVDG 766
QY 672 SPIARPLSFTFPDQVATYG-----ISSQFLGRGIMVSPVLQ-----PGSS-----IYN 715
DB 767 MPLARSMLLTDTEDTTFNFSQKFLDQYMWAGDDILVAPILHSRNEVPGENRVDYPLFHF 826
QY 716 AYSR-----GNWV---SLNSTSSVSVSAGTYVSLSAPPDH-----INVHI 754
DB 827 TWYPSNLRPWDDQGVALGNPVEGGSVINYTARI-----VAPEDYNLFNHYVPVYI 876
QY 755 HEGNIV 760
DB 877 REGAII 882

RESULT 7
US-08-633-770A-1
; Sequence 1, Application US/08633770A
; Patent No. 5908760
; GENERAL INFORMATION:
; APPLICANT: Bojsen, Kirsten
; APPLICANT: Yu, Shukun
; APPLICANT: Kragh, Karsten

```

```

; APPLICANT: Christensen, Tove
; APPLICANT: Marcussen, Jan
; TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM A FUNGUS, ITS
; TITLE OF INVENTION: PURIFICATION GENE CLONING AND EXPRESSION IN MICROORGANISMS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/633,770A
; FILING DATE: July 8, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP94/03398
; FILING DATE: OCT-15-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Altman, Daniel E
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: DYOUE.001APC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714-760-0404
; TELEFAX: 714-760-9502
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1066 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-633-770A-1

Query Match 10.3%; Score 498.5; DB 2; Length 1066;
Best Local Similarity 24.2%; Pred. No. 1.4e-35;
Matches 200; Conservative 118; Mismatches 272; Indels 237; Gaps 37;

QY 158 TPFGTIYKSTHVDLFDATIP-----SNPTTF-----LIYKQDYL 193
DB 152 SFRIQVRLT--PLKDPYIPNVAAAEARVSDKVVWQTSPTFRKMLHPQHKLKDTV 209
QY 194 LQLSSSLPAQAQAHLYGLGE-----HTKPTF-----QLAHNQILTLWNADIASFNRLD 240
DB 210 LDI---VKPGHGYVGMCGEIEFMKEPTFMVNFDMQYQVY---AQGALDSRE- 261
QY 241 NLYGSHPFYMDVRSSPMVSGTHGVFLNMGDMVEYTGDRITYKVG---GIIDLY-IFA 296
DB 262 PLYHSDPPFLVDNSNPEHKNITATFIDNYSQIADFGKTSNGYIKLGRYGGIDCYGISA 321
QY 297 GRTPEMVLQDVKTLIGRAPMPYAFGPHQCRGWYRDVNEIETVVDKYAEARIPLEVMWT 356
DB 322 DTVPDIV-RLYTGLVGRSKLPRYLIGHAQACYGQESDLYSVVQYRDCKFPLDGIHV 380
QY 357 DIDYMDAFKDFLDPVHFPLDKMQQFVTKLHRNGORYVPIIDPGINTNK---SYGTFRIG 413
DB 381 DVDVDQGFRTTNPITFPNPK--EMFTNLRNNGIKCSTNITPVISINNREGGYSTLLEG 438
QY 414 MQSNVFIK-----RNGN-----425
DB 439 VDKXYFIMDDRYTEGTSQNAKDVRYMYGGGNKVEVDPNVNGRDPFKNDYDFPANFSK 498
QY 426 --PYLGSWMPG---PVYYPDFLPPAARSFWDDEIKRFRDILPIDGWIWMNEASNFIT 478
DB 499 QPYTHGGSYGVGNGSAGFYPLDNRKEVRIWMGMQYKFLD-MGLEFWQDMT-----550
QY 479 SAPTPGSTLDNPYPYKINNSGGRVPINSKTIPTATAMHYGNVTEYNAHN-----L 526

```



US-09-275-608-4  
; Sequence 4, Application US/09275608  
; Patent No. 6541237  
; GENERAL INFORMATION:  
; APPLICANT: YU, Shukun  
; APPLICANT: JSEN, Kirsten  
; APPLICANT: MARCUSSEN, Jan  
; TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE AND  
; TITLE OF INVENTION: ITS USE IN THE PRODUCTION OF 1,5-ANHYDROFRUC  
; TITLE OF INVENTION: TOSE  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson & Bear  
; STREET: 620 Newport Center Drive 16th Floor  
; CITY: Newport Beach  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows  
; SOFTWARE: FastSeq for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/275,608  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP95/02172  
; FILING DATE: 06-JUN-1995  
; APPLICATION NUMBER: 08/836,156  
; FILING DATE: 15-APR-1997  
; APPLICATION NUMBER: PCT/EP94/03397  
; FILING DATE: 15-OCT-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Altman, Daniel E  
; REGISTRATION NUMBER: 34,115  
; REFERENCE/DOCKET NUMBER: DY09.001C1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 714-760-0404  
; TELEFAX: 714-760-9502  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 570 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-275-608-4

Query Match 6.9% Score 336; DB 4; Length 570;  
Best Local Similarity 24.3%; Pred. No. 2.le-21;  
Matches 140; Conservative 81; Mismatches 210; Indels 146; Gaps 25;

Qy 226 LTNADIASFNRLNLYGSH-----PFYMDVRSSPMV-----GS-----THGVFL 267  
Db 1 MTNYVNLNYPQDLPPGHSDPDYVIMYF---AAPWIAHGYRGTSQVSYGWF 57  
Qy 268 N-----SNGHDVEYTGDR-ITY--KVIGGIIDLYIF-AGRTPEWLDQYTKL-G----- 312  
Db 58 NVSQSYTNTGDDAWAGQKDLAYMGAQCQPFDOHFVEAGDGLDVVTAFSYLGQKEYENQ 117  
Qy 313 ----RPAPMPYWAGFGHCQWGRDV-----NETETVVDKYAEARIPLEYMWT 357  
Db 118 GLNRSAMPKYVEGFGQVFGATSLRDNLNLAGENVVLEEIVGYNQNVFPBGLAVD 177  
Qy 358 IDYMDAFKDFTLDPVHFPLDKMQQFVTKLHRNGORYVPIIDPGINTNKSYSYGTFI--RGM- 414  
Db 178 VDMQDDLRFVITTRPAFTANK-----VGEEDPNKSVPEWAHDGLV 220  
Qy 415 -QSNV--FIKRNPNP-----LGSVWPGP----- 435

Db 221 CQNTVTCFLKNEKNPYEVNQSLREKQLYTKSDSLDNIDFGTTPDGPSDAYIGHLDYGGGV 280  
Qy 436 ---VYYPDFLDPAAARSFWVDEIKRFRDILPDIQWIDMNEASNFITSAPTP-----GSTL 487  
Db 281 ECDALFPDWGHPDVNAQWGDNYKLFISI-GLDFWQDMTVFAMPHRLGLDPVGTNSGETA 339  
Qy 488 DNPYKINNSGSRVPINS--KTIPATAMHYGN-----VTEYNAHNLYGFLESQATREAL 539  
Db 340 PGWPNKDPSPNGRYNWKSYHPQVLVTDWRYDDYGRDPIVTQORNLH-AYTLCES-TRREGI 397  
Qy 540 VRPAT-----RGPFLLSRSTFAGSGKYTAHTWTGNAARWDDLOYSIPTMLNFGLFGRPMI 594  
Db 398 VGNADSLTKFRSYYIIISRGYIGNQHPGGMVGVGNSSTEDYLAAMVINVINMNSGVPLV 457  
Qy 595 GADICGFAE-----STTEELCCRWIQLCAFYPPFGRDHSAR-----DTTHQELLYLVESV 642  
Db 458 GSDIGGFTEHDKRNPCTPDLMRFFVQAGCLLPWFNRHYDRWIESKKHGKKNYQELIYR 517  
Qy 643 AASARTVLGLRYELPPYYTYTLMYDANLRGSPARPLS 679  
Db 518 LDALRSFVELRYRQEVLYTAMYQNALNGKPIKTVS 554

RESULT 10  
US-08-633-768A-2  
; Sequence 2, Application US/08633768A  
; Patent No. 6013504  
; GENERAL INFORMATION:  
; APPLICANT: YU, SHUKUN  
; APPLICANT: BOUSEN, KIRSTEN  
; APPLICANT: KRAGH, KARSTEN  
; APPLICANT: BOJGO, MAJA  
; APPLICANT: NIELSEN, JOHN  
; APPLICANT: MARCUSSEN, JAN  
; TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM  
; TITLE OF INVENTION: A FUNGUS INFECTED ALGAE, ITS PURIFICATION  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson & Bear  
; STREET: 620 Newport Center Drive 16th Floor  
; CITY: Newport Beach  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/633,768A  
; FILING DATE: 02-JUL-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 9321301.5  
; FILING DATE: 15-OCT-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Altman, Daniel E  
; REGISTRATION NUMBER: 34,115  
; REFERENCE/DOCKET NUMBER: DY09.001APC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 714-760-0404  
; TELEFAX: 714-760-9502  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1091 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-633-768A-2



QY 461 LPIDGIWIDM-----NEASNFIPTAPTGSTLDNPPYKINNSGGRVPINSK-----IP 509  
DB 596 -GLDFVWQDVTVPAMPHKIGDDINVKPD-----GNWPNADPSNGO--YNNKTYHPQVL 647  
QY 510 ATAMHYGN-----VTEYNAHNLVGFLESQATREALVRPAT-----RGPFLLSRSTFAGS 558  
DB 648 VTDMRYENHGREPMVTQRNIH-AYTLCES--TRKEGIVENACTLTKFRRSYIISRGYIGN 705  
QY 559 GKTAHWTGDNARWDDLOYSIPTMLNFGFLGMPMIGADICGFAESTTE-----EJC 610  
DB 706 QHFGGVMVGDNSTSYIGMMIANINNMNSCULPVGSDIGGTSYDNENQRTPTGGLM 765  
QY 611 CRWILQAGFYPPSRDHSAR-----DTHQELVMSVAASARTVLGRYELLPPYYTL 663  
DB 766 VRYVQAGCLLPWFENHYDRWIESKDHGKDYQELMYMPNEMDILRKVFERYRWGEVLYTA 825  
QY 664 MYDANLRGSPDIARPLSFTFPDDVATYGISOFLI-----GRGIMVSPVLOFGSSIVNAYS 719  
DB 826 XYQNAAFGKPIKAASKYNNDSNVRRAQNDHFLGGHDGYRILCAPVMWENSTERELYLP 895  
QY 720 -RCMWV-----SLSNYSVSVSAGTYV-SLSAPPDHNHVIHEGNIV-----AMQGEAMTT 769  
DB 886 VLTQWYKFGDPDTKPLEGAMNGDRIYNTVPQSEPIPVREGALDPRYLNGENKSL 945  
QY 770 QA-ARSTPFHLLVMSDHVASTGELFLDNGIEMDIOGPGGKWTLVRFAB--SGINNLT 826  
DB 946 NTYTDEDPVFEVFPPLGNRRADGMCYLDGCVTTNAEDNGKFSVVKVAEODGTETITP 1005  
QY 827 SSE 829  
DB 1006 TND 1008

## RESULT 12

US-09-634-238-331  
; Sequence 331, Application US/09634238  
; Patent No. 6544772  
; GENERAL INFORMATION:  
; APPLICANT: Glenn, Matthew  
; APPLICANT: Havukkala, Ilkka J.  
; APPLICANT: Bloksberg, Leonard, N.  
; APPLICANT: Lubbers, Mark W.  
; APPLICANT: Dekker, James  
; APPLICANT: Christenson, Anna C.  
; APPLICANT: Holland, Ross  
; APPLICANT: O'Toole, Paul W.  
; APPLICANT: Reid, Julian R.  
; APPLICANT: Coolbear, Timothy  
; TITLE OF INVENTION: Polynucleotides, materials incorporating  
; TITLE OF INVENTION: them and methods for using them.  
; FILE REFERENCE: 11000.104301  
; CURRENT APPLICATION NUMBER: US/09/634,238  
; CURRENT FILING DATE: 2000-08-08  
; NUMBER OF SEQ ID NOS: 422  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 331  
; LENGTH: 390  
; TYPE: PRT  
; ORGANISM: Lactobacillus rhamnosus  
US-09-634-238-331

Query Match 3.8%; Score 182.5; DB 4; Length 390;  
Best Local Similarity 23.4%; Pred. No. 7.2e-08;  
Matches 70; Conservative 46; Mismatches 146; Indels 37; Gaps 11;  
QY 451 VDEIKRFRDILP-----IDGIWIDNEASNFITSAPTGSTLDNPPYKINNSGGRVPINSK 506  
DB 106 LQNLKSPADYADQHGVTGLWTQCN-----LSPVDP-----NPKPDRDFAXEVALGVK 155  
QY 507 TIPATAMHYGNVTEYNAHNLVGF-LESQATREAL-----VRPATRGPFLLSRSTFAGSKYT 562  
DB 156 ALKTDVAWVGSG-----YSFGLDGLAKADAMMTQVKGDSURPFAITLDGMAGTQRYA 207

QY 563 AHWTGD-NAARWDDLOYSIPTMLNFGFLGMPMIGADICGFAESTTEELCCRWIOLGAFYP 621  
DB 208 GWTGDDQTGGQWEYIRPHIPTYIGTGLSGQPVGSDMDGIFGGGNPIVNTDRFQWKATP 267  
QY 622 FSRHSARDTHQELVLM-ESVAASARTVLGRYELLPPYYTLMYDANLRGSPDIARPLSF 680  
DB 268 IQLANDMGNGANPKTPFPDFDOQTAINRAYNKQKTMALMPYNTYTASQAQSVFDGKPMVRGLFL 327  
QY 681 TFPDDVATYG--ISSOFLIGRGIMVSPVLO-----PGSSIVNA-YSPRGNWVLSLNT 730  
DB 328 DYPNIPZAYTDLVKYELWGNFLVAPIYQNTAADEKGNVNRNGIYLPDKQOVWIDYYT 386  
RESULT 13  
US-09-437-054A-10  
; Sequence 10, Application US/09437054A  
; Patent No. 6316698  
; GENERAL INFORMATION:  
; APPLICANT: Allen, Stephen M.  
; APPLICANT: Kinney, Anthony J.  
; TITLE OF INVENTION: Plant Alpha-Glucosidase II Homologs  
; FILE REFERENCE: BB1273 US NA  
; CURRENT APPLICATION NUMBER: US/09/437,054A  
; CURRENT FILING DATE: 2001-05-14  
; PRIOR APPLICATION NUMBER: 60/107,909  
; PRIOR FILING DATE: 1998-No. 6316698ember-10  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 10  
; LENGTH: 150  
; TYPE: PRT  
; ORGANISM: Triticum aestivum  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (73)..(74)  
; NAME/KEY: UNSURE  
; LOCATION: (102)  
; NAME/KEY: UNSURE  
; LOCATION: (123)  
; NAME/KEY: UNSURE  
; LOCATION: (141)  
; NAME/KEY: UNSURE  
; LOCATION: (143)  
US-09-437-054A-10  
Query Match 3.6%; Score 175.5; DB 4; Length 150;  
Best Local Similarity 30.1%; Pred. No. 6.5e-08;  
Matches 46; Conservative 24; Mismatches 70; Indels 13; Gaps 5;  
QY 652 LRYELLPPYYTYIMYDANLRGSPDIARPLSFTFPDDVATYGISOFLIGRGIMVSPVLOPGS 711  
DB 4 MRYSLPPYYSLFQEASVTGVPVNGRPLWLEFPDDKRETYNNGEAFWVGPSSILAQGIYEEGQ 63  
QY 712 SIYNAYSFRGN-WVSLN---YTSSVSVSAGTYVLSAPPDHNHVIHEGNIVAMQG--E 765  
DB 64 KSVSVYLFEXXLMYDLRNGSPYKGSV-----HKLQVSEDSIPXFSQSGTIVPRKDR 117  
QY 766 AMTQAAKSTPFLHVVMSDHVASTGELFLDNG 798  
DB 118 RILTOXVNDIS-YTLVIGLNNWLVXKXTIVDDG 149

## RESULT 14

US-08-306-546C-2  
; Sequence 2, Application US/08306546C  
; Patent No. 5605797  
; GENERAL INFORMATION:  
; APPLICANT: Friderici, Karen  
; APPLICANT: Jones, Margaret  
; APPLICANT: Chen, Hong  
; APPLICANT: Cavanagh, Kevin  
; TITLE OF INVENTION: Bovine Beta-Mannosidase Gene and Methods



1 TITLE OF INVENTION: of Use  
2 NUMBER OF SEQUENCES: 28  
3 CORRESPONDENCE ADDRESS:  
4 ADDRESSEE: Harness, Dickey & Pierce, P.L.C.  
5 STREET: P.O. Box 828  
6 CITY: Bloomfield Hills  
7 STATE: Michigan  
8 COUNTRY: USA  
9 ZIP: 48303  
10 COMPUTER READABLE FORM:  
11 MEDIUM TYPE: Floppy disk  
12 COMPUTER: IBM PC compatible  
13 OPERATING SYSTEM: PC-DOS/MS-DOS  
14 SOFTWARE: PatentIn Release #1.0, Version #1.25  
15 CURRENT APPLICATION DATA:  
16 APPLICATION NUMBER: US/08/306,546C  
17 FILING DATE: September 15, 1994  
18 CLASSIFICATION: 435  
19 ATTORNEY/AGENT INFORMATION:  
20 NAME: Smith, DeAnn F.  
21 REGISTRATION NUMBER: 36,683  
22 REFERENCE/DOCKET NUMBER: 6550-00003  
23 TELECOMMUNICATION INFORMATION:  
24 TELEPHONE: (810)641-1600  
25 TELEFAX: (810)641-0270  
26 INFORMATION FOR SEQ ID NO: 2:  
27 SEQUENCE CHARACTERISTICS:  
28 LENGTH: 879 amino acids  
29 TYPE: amino acid  
30 TOPOLOGY: linear  
31 MOLECULE TYPE: protein  
32 US-08-306-546C-2

Query Match 2.5%; Score 119.5; DB 1; Length 879;  
Best Local Similarity 19.3%; Pred. No. 0.13; 292; Indels 327; Gaps 51;  
Matches 175; Conservative 111; Mismatches 292; Indels 327; Gaps 51;

QY 61 TALLQLIRNSPVYGPDIHLSFTASFEEDTLR-----IRFTD-----ANNRR-----W 104  
DB 105 TVAVVLNSVPICKTDMFRY--SFDTHTVXAVNIIEVRFQSPVYVYANQSERHTAYW 162  
QY 105 EIPNEVLPRPPSPPLSLSLQHLKPIPNQ-----PTTVLSHPHSLAFT 153  
DB 163 -----VPNCPPPVQDGECHVNFIRKMQCSFGWDMGSPFPQGIWKDVRIE-AYN 211  
QY 154 LPHTPPGFT-IYR-----KSTHVLFDATPIPSNPTTFLIYKQOYLQSSSLPA 202  
DB 212 VCHLNYPMTPIYDNYMKTNWLNKIESSFDVV--SSKLVSGEAIVAI-PELNIQOTNNIEL 268  
QY 203 QOAHLYGLGEHTKPTFQLAHNQIL--TLW-----NADI-----ASPNRDLNLY 243  
DB 269 QH-----GERTVELFKIDKAIIVETWMPHGHGNQGYNWSVIFELDGLRFEKSAKVY 322  
QY 244 -----GSHP-----FMDVRSSPMVSGTHGVFLNNGMDVEYTGDRITYKVIIGI 289  
DB 323 FRTVELVEEPIONSPGLSFYFKINGLP-----IFLKGSNWIPADSFQDRVTSAMRLLL 375  
QY 290 IDLYIFAGRTPEMYLDQYTKLIGRPAPMPYATGFGHCRMGVDRVNEIETVVDKYEARI 349  
DB 376 LQSVDDANMALRV-----WGGGVYB-----QDEFYELCD 405  
QY 350 PLEVW-WTDIDYMDAF-----KDFTLDPVHFPDLKMQOFTVTLKLRNGQRYVPILD-PCINT 403  
DB 406 ELGIMIWDPMFACALYPTDKDF-----MDSVREEVTHQVRRLKSHPSLIITWSGNE 457  
QY 404 NKS-----YGTFRMGQSNVFIKRNPNYLGSVWPGVTVYDPFLDPAARSFWVDIKRF 457  
DB 458 NEAALMMGWYDT-----KPGYLQTYIKDYVTVLVKNI 489  
QY 458 RDILPIDGIMDMNEASNFITSPTPGST-----LDNPPYKINNNGRVPPI-----503  
DB 490 RTIV-----LEGQTRFFITSSTPTGAKTIAEGWLSFNPYDLNY--GDVHFYDYVSDCW 541

QY 504 NSKTIPATAMHYGNVTEYNNAHNLGYF--LESQATREALVRPATRGPFLLSRSTFAGSGKY 561  
DB 542 NWRTPFKARF-----VSEGYQSWFSFSTLEKVSSEEDW---SYRSSFALHROHLINGNE 594  
QY 562 TAHW-----TODNAARWDDIQQYSIPTTMLNFGLEFMPMIGADICGFAESTTTEELCCR 612  
DB 595 MLHQIELHFKLPNSTDQLRRFKDTLYLTQVMQ-----AQCVKTEFE-----635  
QY 613 WQLGAFYPPSRDH--SARDTHQELY-----LMESVAASARTVLGLRYELLPPYYTLMY 665  
DB 636 -----FYRRSRSEIVNGKGTMGALYWLNDIWOAPSWSS-LEYGGKWKMLHYFARHFF 688  
QY 666 DANLRGSPRIARPLSFTFPDDVATVIGS-----SOFLI 697  
DB 689 -----APLL-FVGFEKXDMLFIVGASHLHSDQOMLTVRVHTWSSLELVCSSEINPFI 741  
QY 698 GRG-----IMVSPVLQ-----PG-----SSIVNAY-SPRGNWVLSNY--TSSVSVSAGTYV 741  
DB 742 KAGESVLLYTKVPPELLKGCPCGCTRCQSCVSYLSTDGELISPINYHFLSLKNAKG---798  
QY 742 SLSAPPDHINVIHEGNI---VAMQGEAMTTQAARS--TPHLLVWMSDHYASTGELFLD 796  
DB 799 -----LHKANITATISQOQDTPFVPLKTSVAVPFVWL-----DVGSIPGRPSD 841  
QY 797 NGIEM 801  
DB 842 NGFLM 846

## RESULT 15

US-08-530-524A-2  
; Sequence 2, Application US/08530524A  
; Patent No. 5837836  
; GENERAL INFORMATION:  
; APPLICANT: Friderici, Karen  
; APPLICANT: Jones, Margaret  
; APPLICANT: Chen, Hong  
; APPLICANT: Cavanagh, Kevin  
; TITLE OF INVENTION: Bovine Beta-Mannosidase Gene and Methods  
; TITLE OF INVENTION: Of Use  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.  
; STREET: P.O. Box 828  
; CITY: Bloomfield Hills  
; STATE: Michigan  
; COUNTRY: USA  
; ZIP: 48303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/530,524A  
; FILING DATE: September 19, 1995  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Smith, DeAnn F.  
; REGISTRATION NUMBER: 36,683  
; REFERENCE/DOCKET NUMBER: 6550-00003DVA  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (810)641-1600  
; TELEFAX: (810)641-0270  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 879 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-530-524A-2

Query Match 2.5%; Score 119.5; DB 2; Length 879;

Search completed: October 27, 2003, 10:26:41  
Job time : 22.5211 secs







ID ABB09152 standard; Protein; 901 AA.  
 XX  
 AC ABB09152;  
 XX  
 XX 01-JUL-2002 (first entry)  
 XX  
 DE Buckwheat alpha-glucosidase protein SEQ ID NO:2.  
 XX  
 DE Buckwheat; alpha-glucosidase; enzyme; Fagopyrum esculentum Moeench;  
 KW plant.  
 KW  
 XX Fagopyrum esculentum.  
 XX  
 XX Key Location/Qualifiers  
 FT Peptide 1..31  
 FT /label= signal  
 FT Protein 32..901  
 FT /label= alpha\_glucosidase  
 XX  
 XX JP2002065272-A.  
 XX  
 XX 05-MAR-2002.  
 XX  
 XX 31-AUG-2000; 2000JP-0262102.  
 XX  
 XX 31-AUG-2000; 2000JP-0262102.  
 XX  
 XX (NISO ) NIPPON SHOKUHIN KAKO KK.  
 XX  
 XX WPI; 2002-346763/38.  
 DR  
 DR N-PSDB; ABL51447.  
 XX  
 XX A buckwheat-derived alpha-glucosidase gene -  
 XX  
 XX Claim 1; Page 9-10; 21pp; Japanese.  
 XX  
 XX The present sequence represents an alpha-glucosidase isolated from  
 CC buckwheat (Fagopyrum esculentum Moeench). The alpha-glucosidase gene  
 CC can be used for the preparation of alpha-glucosidase derived from  
 CC buckwheat.  
 XX  
 SQ Sequence 901 AA;  
 Query Match 58.1%; Score 2824; DB 23; Length 901;  
 Best Local Similarity 60.2%; Pred. No. 1.2e-219;  
 Matches 542; Conservative 136; Mismatches 200; Indels 22; Gaps 10;  
 QY 14 LAVVPLVLCWVEGATTSKNDQGEAIGYGYQVNAKVDNSTGKSLTALLQIRNSPVY 73  
 DB 14 LLLAATLLFCSLF---VVSSED---EVVGYGYRVRAKVDSSS-NLTAFKLINASSLY 66  
 QY 74 GPDIFLSFTASFEEDDTLRIRFTDANNRWEIPNEVLPRPPPPPPPLSSLOHLPKPI 133  
 DB 67 GQDIPNLTFTATFEKQYRLRIRITDAEKRWEPNEVLHRDGSNGHGHQPLDSR----- 119  
 QY 134 PQNQPTTTLVLSHPSDLAFLFHTTPFGFTIYRKSTHDVLFDA-TPIPSNPTFL-YKQ 192  
 DB 120 PTPPSAAVLTHPNSDLIRLHDTPFGFVTRSRSTNDVLFDTRESADPTDPGLVFKQ 179  
 QY 193 YLQLSSLPAAQAHYGLGHEHTKPTFOLAHNQLTLNADIASFNRLNLYGSHPFYMDV 252  
 DB 180 YIQLSSLPADRSNLYGIGHEHTKPTFLARNQTLTLNADIASVNVDLNLYGSHPFYLDV 239  
 QY 253 RSSPMVSTHGVLNMGMDVEVTGRTIYKVIIGGIIDLYIFAGRTPEMVLQDYTKLIG 312  
 DB 240 RAP--LGTSGVLLNMGMDVEVTGDKITYKVIIGGIIDLYIFEGPTFEVQOQYTELIG 297  
 QY 313 RPAPMPYWFAGFHOCRWGYRDVNEIETWVDKAEARIPLEVMWMTDIDYMDAFKDTLDPV 372  
 DB 298 RPAPMPYWSFGFHOCRWGYEYRNSVENVVAKYSTMRIPLEAMWTDIDYMEANKDFTVDPV 357  
 QY 373 HFPLDKWQFVTKLHRNGQRYVPILDPGINTN-KSYGTFFIRGMSNVFIKRNNGNVLGVSV 431  
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

DB 358 NFPLDKWQFVYNKLIHQKQYVAILDPGININTTTTGTFFORAMKADIFIKRQGEPEYQGEV 417  
 QY 432 WPGVYYPDFLDPAARSFWDEIKRFRDILPIDGIWIDMNEASNFITSAPTFGSTLDNPP 491  
 DB 418 WPGVYYPDFLDNPKTTFIWFSEIQTFNFALPDGLWIDMNEVSNFISPPIDPSLDNPP 477  
 XX  
 QY 492 YKINNSGRVPINSKTIPTATAMHYGNVTYNAHNLGYFLESQATREALVRPATRPGPFLS 551  
 DB 478 YVINNSGRRRINEKTIPTVSSVHYGNVSDYVNRHLYGILEATATNVALKVTKORPPVLS 537  
 QY 552 RSTFAGSGKYTAHMTGDNAARDLQYSIPTMLNFGFMPEMIGADICGFAESTTEELCC 611  
 DB 538 RSTFIGSGKYTAHMTGDNAATWNNAAASIPITLDFGLFGIPMIGADICGFPAYTTTEELCR 597  
 QY 612 RWIQLGARYPFESRDHSARDTHOBYLYLWESVAASARTVLGLRYELLPPYYTYLMDANLGR 671  
 DB 598 RWIQLGARYPFESRDHSNLSSEPQLTQGSVTSERKVLGLRYELLPPYYTYLMDANLGR 657  
 QY 672 SPIARPLSFTFPDDVATYGISSQFLIGRIMVSPVLOPGSSIVNAYSPRGNWVLSNYTS 731  
 DB 658 TPIARPLSFTFPDDVATYGISSQFLIGRIMVSPVLOPGSSIVNAYSPRGNWVLSNYTS 717  
 QY 732 SVYSVA-QTYVLSAPPDHIHNVHIEGNIVAMQGEAMTQAARSTPPELLVMSD--HVA 788  
 DB 718 TVSSPANGSFVTLDAPLKEINNVHVRREGSILALQGEAMTTREARNTPELVVVISDSGNGS 777  
 QY 789 STGELFLDNGIEMDIGGPGKMTLVRFPAB-SGINNLTISSEVNVNRYGAMSORVMDKIT 847  
 DB 778 SIGSVFLDNGVDIEMDGGGRWSLVTFSGLVGNKNKVTITSSVNVNGFALSQGWKISKVT 837  
 QY 848 ILGLKRRVKIKEYTVQKDAGAKVKGLGRRTSSNQGGFFVSVISDLRQLVGOAFKLELE 907  
 DB 838 ILGLSRGQVKGTYMLSIGRVVTTISAKGARKCKGTGKFDVVEIPNLSLLVGRNFKLDIQ 897  
 RESULT 5  
 AAM59040  
 ID AAM59040 standard; Protein; 877 AA.  
 XX  
 AC AAM59040;  
 XX  
 DT 07-AUG-1998 (first entry)  
 XX  
 DE Barley alpha-glucosidase protein.  
 XX  
 KW Alpha-glucosidase; barley; starch grain hydrolysis; alpha-amylase;  
 KW glucoamylase; industry; germplasm; hydrolytic enzyme.  
 XX  
 OS Hordeum vulgare.  
 XX  
 PN US5763252-A.  
 XX  
 PD 09-JUN-1998.  
 XX  
 PF 28-APR-1995; 95US-0430925.  
 XX  
 PR 28-APR-1995; 95US-0430925.  
 XX  
 PA (WISC ) WISCONSIN ALUMNI RES FOUND.  
 XX  
 PI Skadsen RW, Tibbot BK;  
 XX  
 DR WPI; 1998-347329/30.  
 XX  
 DR N-PSDB; AAV11736.  
 XX  
 PT DNA encoding barley alpha-glucosidase protein - useful for producing  
 PT recombinant protein to increase rate of starch grain hydrolysis when  
 XX used with alpha amylase  
 PS Claim 1; Col 21-26; 19pp; English.  
 XX  
 CC This sequence represents a novel barley alpha-glucosidase protein.  
 CC Recombinant alpha-glucosidase can be used to increase the rate of starch











```
QY 60 LTALLQLI-RNSPVYGPDIHFLSFSTASFEEDDTLIRF-----TDANRRWEIPNEVLPRP 114
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 78 FTASQLAGRPNVYGTDESILTSVEYQSDRLNIQILPHTVDSINASWYFLENL--- 134
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 115 PPPSPPLSSLOHLKPFPQNPQTTTTLVSHPSDIAFTLFTHTTTPGFTTYRKSTHDLVF 174
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 135 ----VPRPKASLN-----ASVSQDLFVSWSNPSFNFVKIRKATGDALE 175
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 175 DATPISNPTTFLIYKDQYLQSSLPAAQAHLYGLGHEKTPFOLAHNOILTLWNADIA 234
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 176 -----STEGTVLVYENQFIEFTALP-EYNYLGLGHEI-TQFLORNANLTIYPSDDG 227
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 235 SFNRDLNYGSHPPYMDVR-----SSPMVSGTHGVFLNLSNGMD 273
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 228 T-PIDQNLVYGHPPYLDTRYKGRQNGSYIPVKSSEADASQDYISLSHGVLFRNSHGLE 286
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 274 VEYTGDRITYKVIIGIIDLIFAGRTPEMLDQY-TKLIGRPAPMPYAFGPHQCRWGYR 332
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 287 ILLRSQKLIRWTLGGGIDLTIFYSGPAPADVTRQYLTVGLPAMQYNTLGFHQCRWGN 346
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 333 DVNEIETWVKYAEARIPLEVMWTDIDYMDAFKDFTLDPVHFPLDKMQQFVTKLHRNGOR 392
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 347 NWSLDADVAVNFEKFEIPLEYITWIDYMHGYRNFNDQHRFSYSEGDEFSLKHSRGY 406
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 393 YVPILDPGI-----NTNKS YGTGFIHQMSNVFIKR-NGNPYLGSVWPGVYYPDFLPA 445
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 407 YVPIVDAALYIPNPENASDAYATYDRGAADVFLKNPDGSLYIGAVWPGYTVFPDWHHPK 466
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 446 ARSFWDIEIKRFRDILPIDGIWIDNEASNP-----PTPG-STLDNPPYKINN 496
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 467 AVDFMANELWIWSKKYAFDGVWYDMSEVSSFCVSGCGTGNLTLPNPAHPSFLLEPGDII 526
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 477 -----ITSA-----TPPG-STLDNPPYKINN 496
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 527 YDYPEARNITNATAEASASAGASSQAATATTSTSVSYLRTTTPGVNRVHEPPYVNH 586
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 497 SGRVPINSKTIPTAMHYGNVTYNAHNLVGFLESQATREAL--VRPATRGPFLLSRST 554
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 587 DQEGHDLNVHVASPNATHVDGVEYDVHGLYGHQGLNATYQGLLEVWHSKRRPFIIGRST 646
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 555 FAGSGKYTAHTGNAARWDDLOYSIPTMLNFGIFGPMIGADICGFAESTTRELCCRWI 614
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 647 FAGSGKWAGHWGDDNYKSMWYSISQALSFLGIFPMFGADTCGFNGSDBELCNRMW 706
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 615 QLGAFYFSDHRSARDTHOELYLWESVAASARTVLGLRYELLYYYTLMVDANLRGSP 674
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 707 QLSAFFFFYRNHNLSTIPQEPYKASVIEATKSAKIRVAILPYFTLFDLAHTTGSTV 766
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 675 ARPLSFTPDVATYGISQFLIGRMVSPVLOGSSIVNAYSP---RG-----NMVSL 727
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 767 MRALSWEFPNDPTLAAVETQPMVGPALMVVPLEVNTVKVPPGPGVGHGEVWYDM--- 822
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 728 NYT-SSVSVSAGTVLSLAPDHNHVIHHEGNIAMQGEAMTTOARSTPFLHIVMSOH 786
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 823 -YTOAAVDKRGVNTTISAPLGHIPVYRGNNILPMQEPALTITREAKQTPWALLAUGSN 881
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 787 VASTGEFLONGIEMDIGGPGKKTTLVRFASGGINNLTISSSE-VNRYGAMSQ---RW- 841
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 882 GTASQLYLDG-----ESYIPNATLHVDTASRSS:RSSAQCGRWK 922
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 842 ---VMDKITLGLKRRVKIKETVOKDAGAKIKVG 873
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 923 ERNPLANVTVLG-----VNKPSAVTLNG 946
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
AAR42214
ID AAR42214 standard; Protein: 985 AA.
XX AC AAR42214;
XX DT 19-MAY-1994 (first entry)
XX
```

```
DE Aspergillus niger glucosyltransferase.
XX
KW GTase; glucoamylase; GHase; recombinant production; truncated;
KW N-terminal deletion; C-terminal deletion.
OS Aspergillus niger #499.
PN JP05268953-A.
XX
PD 19-OCT-1993.
XX
PF 27-MAR-1992; 92JP-0101761.
XX
PR 27-MAR-1992; 92JP-0101761.
PA (AMAN) AMANO PHARM KK.
PA (UOZU) UOZUMI T.
XX
DR W21; 1993-364284/46.
DR N-PSDB; AAQ50982.
XX
PT Glucoamylase prepn. with reduced glucosyl-transferase activity -
PT using microorganism transformed using plasmids including DNA
PT fragments having deletion of N-end or C-end of Aspergillus
PT glucosyl-transferase
XX
PS Disclosure; Page 6-9; l1pp; Japanese.
XX
CC DNA encoding truncated variants of Aspergillus glucosyltransferase
CC is claimed. Deletion from the N- or C-terminal reduces GHase
CC activity of the protein. Specifically, glucoamylase with reduced
CC GHase activity can be prepared. The full-length GHase structural
CC gene was isolated as a 4.3kb SphI fragment from chromosomal DNA of
CC A.niger #499.
XX
SQ Sequence 985 AA;
Query Match 27.6%; Score 1339.5; DB 14; Length 985;
Best Local Similarity 32.6%; Pred. No. 3.3e-99;
Matches 324; Conservative 156; Mismatches 306; Indels 209; Gaps 29;
QY 7 PRYICPTLAVLPLVLCMVVEGATTSKN-----DNQGEAIGYGYQVNAKVDNSTGKS 59
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 60 LTALLQLI-RNSPVYGPDIHFLSFSTASFEEDDTLIRF-----TDANRRWEIPNEVLPRP 114
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 78 FTASQLAGRPNVYGTDESILTSVEYQSDRLNIQILPHTVDSINASWYFLENL--- 134
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 115 PPPSPPLSSLOHLKPFPQNPQTTTTLVSHPSDIAFTLFTHTTTPGFTTYRKSTHDLVF 174
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 135 ----VPRPKASLN-----ASVSQDLFVSWSNPSFNFVKIRKATGDALE 175
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 175 DATPISNPTTFLIYKDQYLQSSLPAAQAHLYGLGHEKTPFOLAHNOILTLWNADIA 234
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 176 -----STEGTVLVYENQFIEFTALP-EYNYLGLGHEI-TQFLORNANLTIYPSDDG 227
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 235 SFNRDLNYGSHPPYMDVR-----SSPMVSGTHGVFLNLSNGMD 273
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 228 T-PIDQNLVYGHPPYLDTRYKGRQNGSYIPVKSSEADASQDYISLSHGVLFRNSHGLE 286
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 274 VEYTGDRITYKVIIGIIDLIFAGRTPEMLDQY-TKLIGRPAPMPYAFGPHQCRWGYR 332
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 287 ILLRSQKLIRWTLGGGIDLTIFYSGPAPADVTRQYLTVGLPAMQYNTLGFHQCRWGN 346
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 333 DVNEIETWVKYAEARIPLEVMWTDIDYMDAFKDFTLDPVHFPLDKMQQFVTKLHRNGOR 392
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 347 NWSLDADVAVNFEKFEIPLEYITWIDYMHGYRNFNDQHRFSYSEGDEFSLKHSRGY 406
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 393 YVPILDPGI-----NTNKS YGTGFIHQMSNVFIKR-NGNPYLGSVWPGVYYPDFLPA 445
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 407 YVPIVDAALYIPNPENASDAYATYDRGAADVFLKNPDGSLYIGAVWPGYTVFPDWHHPK 466
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```



```

Db 849 PFGILVALDNDGKAQGSGLYLDGGSVDS---LIVSF---SVSDNTLSAS-PSGDY 899
QY 836 AMSQRWMDKITILGLKRRVK-IK-----EYTVOKDAGAIKVKGLGRRSSHNQGGF 886
Db 900 KADQ--PLANVTILGVGHKPKSVKFENAVDFTYKK--STVFTGLDKYT---KDGAP 950

RESULT 13
AAR07575
ID AAR07575 standard; protein; 958 AA.
AC AAR07575;
XX
XX
DT 25-MAR-2003 (updated)
DT 02-FEB-1991 (first entry)
DE Glucoamylase encoded by GAMI gene.
XX
XX Schwanniomycetes yeast cells; expression cassette; glucoamylase;
KW marker gene; regulon; signal peptide; terminator;
KW autonomously replicating sequence.
XX
XX Schwanniomycetes occidentalis.
OS
XX
XX EP394538-A.
PN
XX
XX 31-OCT-1990.
PD
XX
XX 28-APR-1989; 89EP-0107780.
PP
XX
XX 28-APR-1989; 89EP-0107780.
PR
XX
XX (RHEI-) RHEIN BIOTECH GES.
PA
XX
XX Hollenberg C, Strasser A;
PI
XX
XX WPI; 1990-328670/44.
DR
XX
XX N-PSDB; AAQ06389.
XX
XX
XX Transformed Schwanniomycetes yeast cells - contg. an expression
PT cassette contg. regulon, DNA coding for foreign protein and
PT terminator
XX
XX Disclosure; Fig 2B(1-9); 59pp; English.
XX
XX The expression cassette may contain a regulon (R), a signal peptide
CC sequence (S), a foreign protein sequence (P) and a terminator (T)
CC of which R and/or S and/or T are derived from the glucoamylase gene
CC of S. occidentalis.
CC R is pref. a 1.3 kb BamHI-PvuII fragment (bases -1 to -320 of this
CC sequence).
CC S contains all or part of one of the sequences represented
CC in AAQ06383-87.
CC F is e.g. a cellulase, interleukin, insulin-like-growth factor,
CC interferon etc.
CC T is pref. all or part of the terminator of this sequence
CC (bp 2875-3320).
CC GAMI acts as selective marker for the vector carrying the expression
CC cassette
CC R, S, F and T may be modified by insertions, deletions or substitutions
CC such that biological activity is retained or improved.
CC See also AAQ06388.
CC (Updated on 25-MAR-2003 to correct PA field.)
XX
XX Sequence 958 AA;
XX
XX Query Match 27.2%; Score 1321.5; DB 11; Length 958;
XX Best Local Similarity 33.2%; Pred. No. 9e-98;
XX Matches 318; Conservative 151; Mismatches 316; Indels 173; Gaps 32;
XX
XX 34 NDN--QGEAIGYQVKNKAVDNSTGSLTALLQIRNSPVGPDHIFLSTASFEDDT 91
XX
XX

```

```

Db 61 NDSAVDANAAGKYDVLNV---TNTPRGLTGILKLKEATNIYGYDFDYLNLTVYQADTR 117
QY 92 LRIRFTDANNRWEIPNEVLPRPPPPPPPLSSLOHLP-----KPIPONQPTTTVLSHP 146
Db 118 LNVHI-----EPTLDSOVFLPESHVLVVKPLVBDGDAQS--YNYFD 153
QY 147 HSDLAFTPHFTTPFGTIYRKSTHDVFDATPISNPTFTFLIYKDQYLQLSLSSLPAAQAH 206
Db 154 NSDLVFE-YSNTDFSEFVIRSTKEVLEST---KGNP---LVFSNQFIQFNSSLPKNHV- 205
QY 207 LYGLGEH-----TKPTFQLAHNQILTLNADIASFNRDLNLGSHPPFYMDVRSSPMVGS 261
Db 206 ITGLGESIHGLVNEP-----GSVKTLFANDVGO-PIDGNIYGVHPVLDQRYD--TETT 256
QY 262 HGVFLNLSNGMDVEYTGDRITYKVIIGGIIDLYIFAGRTPEMWLDQYTKLIGRPAMPYWA 321
Db 257 HAVYWR TSAIQEVLI GEE SIWRALS GVIDLYFFSGPTPKDAIQQYVKEIGLPAQPYMS 316
QY 322 FGHQCRWGYRDVNEIETVVDKYAEARIPLEVMWTDIDYMAFKDFTLDPVHFFLDKMQQ 381
Db 317 LGYHQCWGYDTIEKLSEVVENFKFNIPLETIWSIDIDYMSYKDFDYDHRFPLDEYRK 376
QY 382 FVTKLHRNGORYVPILD-----PGINTKSYGTFIRGMOSNVFIKR-NGNPYLGSYWP 433
Db 377 FDELUKKNQHYPILDDAAIYVNPNNADINEYQPFHYGNETDVFELKNPDGSLYIGAVWQ 436
QY 434 GPVYYPDFLDPAARSFWVDEIKRFRDILPIDGIWIDMNEASF----- 476
Db 437 -VTLSRFLS-RKHSMDMKVIKMWYELTFPDGIMADMEVSSFCVSGCGTKYFENPAYP 494
QY 477 -----ITSAPTGCSTLD----- 488
Db 495 PFTVGSKATSYPVGFDVSNASEWKSIOSSISATAKTSTSVSSSSSTIDYMNLTAPGKG 554
QY 489 ---NPPYKINNSGGRVPINSKTIPATAMHYGNVTEYNAHNLGYFLESQATREAL--VRPA 543
Db 555 NINYPYAIYNMQGSDLAATHAVSPNATHADGTVEYDIHNLGYLOENATYHALLEVFPN 614
QY 544 TRGPFLLSRSTFAGSGKTYTAHWTDGNAARWDDLOYSIPTMLNFGLFGMFMIGADICGFAE 603
Db 615 KR-PFMISRSTFPFRAGKWTGHMGDNTADWAYAFSIPQAFSMGIAGLPFFGADYVCGFNG 673
QY 604 STEELCCRWIQLGAFYFSDRHSARDTTHOELYLWESVAASARTVLGLRYELLPPYYTL 663
Db 674 NSDSELCSRWMLGSGFFPFYRNHNYLGAIDQEPYVWESVAEATRTSMAIRYLLPPYYTL 733
QY 664 MYDANLRGSPARPLSFTFPDDVATYGISSOFLICRGIMVSPVLOPGSSIVNAYSPRG-- 721
Db 734 LHESHTTGLPIIDRAFSGWFPNDRSLSGVDNQFFVGDLVTVLPVLEFGVDKVKGVFPGAGK 793
QY 722 -----NWVSLSNYTS-SVSVSAGTYVLSLAPDPDHINVHIHEGNIIVAMQGEAMTTOAARST 775
Db 794 BEVYYDW-----YQREVHFQDKGKNETLDAPLGHIPLIRGGNVLPQTQEPGYTVAESRQN 848
QY 776 PHLLVWMSDHSVASTGELFLDNGIEMTDIGGPGKKTLYRFFAESGINNLTTSSEVWNGY 835
Db 849 PFGILVALDNDGKAQGSGLYLDGGSVDS---LIVSF---SVSDNTLSAS-PSGDY 899
QY 836 AMSQRWMDKITILGL---KRRVKIK---EYTVOKDAGAIKVKGLGRRSSHNQGGF 886
Db 900 KADQ--PLANVTILGVGHKPKSVKFENAVNFTYKK--STVFTGLDKYT---KDGAP 950

RESULT 14
AAW15191
ID AAW15191 standard; Protein; 985 AA.
XX
XX AC AAW15191;
XX
XX DT 18-JUN-1997 (first entry)
XX
XX DE Aspergillus oryzae alpha-glucosidase.
XX

```

KW Fungus; fungal; enhancer element; promoter; recombinant protein.  
 XX Aspergillus oryzae.  
 OS JP09009968-A.  
 PN 14-JAN-1997.  
 PD 29-JUN-1995; 95JP-0163579.  
 PF 29-JUN-1995; 95JP-0163579.  
 PR (KOKU-) KOKUZEI CHO CHOCHAN.  
 PA (OZEK-) OZEKI KK.  
 XX WPI; 1997-126425/12.  
 DR N-PSDB; AAT65017.  
 XX Fungal DNA enhancer element - used to transform other host fungus  
 PT e.g Aspergillus oryzae, to produce large quantities of a gene  
 PT product  
 XX Disclosure; Page 14-19; 25pp; Japanese.

XX The promoter region of the alpha-glucosidase (agda) gene of  
 CC Aspergillus oryzae contains two novel enhancer elements: one  
 CC (designated "enhancer-B") corresponds to the consensus sequence  
 CC CGGNATTTA and the other (designated "enhancer-C") is of sequence  
 CC CCAATCAGCGT. By inserting at least one of the enhancer elements  
 CC into a promoter region which is functional in fungi, the activity  
 CC of the promoter is enhanced. Using such improved promoters, a gene  
 CC of interest can be expressed efficiently in transformed fungi.  
 CC The present sequence is encoded by the agda gene.  
 XX Sequence 985 AA;

Query Match 26.9%; Score 1307; DB 18; Length 985;  
 Best Local Similarity 31.8%; Pred. No. 1.4e-96;  
 Matches 312; Conservative 165; Mismatches 321; Indels 184; Gaps 27;  
 QY 28 GATTGKNDQGAIGY-----GYQVNAKVDNSTGSLTALLQLRN-SPVYGPDIHPLS 81  
 DB 46 GANLVANIDDFQAVNAQSCVCGYKASDVK---HSSQGFASLELAGDFPCNVYGTVDVSLT 102  
 QY 82 FTASFEEDDTLRIR---FTDANRRWEI-PNEVLPRPPPPPPPLSSLQHLKXPIPN 136  
 DB 103 LTVEQAKDRINQIVPYFDASNASWILSEELVFRP-----KASON 145  
 QY 137 QPTTTLVSHPHSDLAFTLFTTTPGFTTYRKSTHVDLFDATPISNPTTF:YKDYQLQL 196  
 DB 146 -----ASVQSDFFVVSMSNPSFNFKVIRKATGDLF-----NTKGSTLVYENQPIEF 193  
 QY 197 SSS:PAQQAHLXGLGEHTKPTFOLAHNQLITLWADIASFNRLNLXGSHFFYMDVR--- 253  
 DB 194 VTLLP-EYNLYGLGERMN-QLRLLENANLTUYAADIAD-FIDNNYGHAFYLDTRYK 250  
 QY 254 -----SSPMVSGTHGVFLNLSNGMDVEYTGDRITVKV:GGIIDLVI 295  
 DB 251 VGGQNKSHTVKSSAEAEQEVSVSHGVFLANAHGQEIILLRDKLIWRLTGGSDVITFY 310  
 QY 296 AGRTPEMVLQY-TKLIGRPAMPYWAFFGHQCRMGYRDVNEIETVVDKYAEARIPJEVM 354  
 DB 311 SGFTQAEVTKQYQLSTVGLPAMQYNTLGFHQCRRGYNWSEFEFDVLANFERFEIPEYL 370  
 QY 355 WTDIDYMAFKDFTLDVPHFPLDKMQQFVTKLHRNQYVPIIDPGI-----NTNKSIG 408  
 DB 371 WADIDYMGYRNFNDQHRFSYEEGEKFLNKLHAGGREWPIVDGALYIPNPENASDAYE 430  
 QY 409 TFIQMGQSNVFIKR-NGNPYLGSVWPGFVYPDFLDPARSFWDEIKRFRDILPIDGIW 467  
 DB 431 TYDRGAKDDVFNKFDGSLYIGAVWPGYTVPDWHHPKASDFWANELVTWKNLHYDGVW 490  
 QY 468 IDMEASNF-----ITSA----- 480

DB 491 YDMAEVSFCVSGCGTGNLSMNPAPHPALPGEPGVVYDYPGEGFNITNATEAASASAGA 550  
 QY 481 -----PTPG-STLDNPPYKINSGRVPINSKTIPTATAMEYGNVT 519  
 DB 551 ASQSAASSTTTSAFYLRTTPTPGVNRVDHPYVNHVQPGHDLVSHVSPNSTHSDGVQ 610  
 QY 520 EYNAHNLVGFLESQATREALVRP-ATRGPFLLSRSTFAGSKYTAHWTCGDAARDDLOQY 578  
 DB 611 EYDVHSLYGHQGINATYHGLLVKWNKRPFIARSTFGSGKNAGHWGGDNFSGKMSFF 670  
 QY 579 SIPTMLNFGMIPGADICGFAESTTEELCCRWIQLGAFYPPSRDHSARDTTHOELYL 638  
 DB 671 SISQALQFSLFGIPMGVDTCGFNGNTDEELCNRMQLSAFFPYRNHNVLSAIPQEPYR 730  
 QY 639 WESVAASARTVLGLRYELLPPYYITLMYDANLRGSPARPLSFTPPDDVATYGISQFLIG 698  
 DB 731 WASVIDATKAAMNIRYAILPYFTLFLAHTTGTVMRALAWEPFNDPSLAAGVQTPLVG 790  
 QY 699 RGMVSPULQPGSSIVNAYSP---RG---NWVSLSNYTSVSVSAGTVVLSAPPDHIN 751  
 DB 791 PSVMVPIVLEPOVDTVQGVPGVHGEVWYDYS---QTAVDAKPGVNTTISAPLGHIP 846  
 QY 752 VHIHEGNIVAMQGEAMTTOAARSTPFLHLLVNMDSHVASTGBELFLDNGIEMDGGPGKWT 811  
 DB 847 VVRGGSILPMQOEVALTTTRDARKTPMSLLASLSNGTASGQLYLDGSEV----- 896  
 QY 812 LVRFPAESGINNLTISSEVNRGYAMSORW-----VMDKITILGLKRR---VKIKYTVQK 864  
 DB 897 -----YPEDTSLVDFLASRSTLRASRG-TWKEANPLANVTVLGVTEKPSVTLNGETLSS 951  
 QY 865 D-----AGAIKVKVGLGRRTS 879  
 DB 952 DSVKYNATSHVHLVGGLOKHITA 973

RESULT 15  
 AAW74090  
 ID AAW74090 standard; Protein; 1827 AA.  
 XX AAW74090;  
 AC AAW74090;  
 DT 04-MAY-1999 (first entry)  
 XX Human hSI protein sequence.  
 DE Gastro-intestinal transport receptor; binding protein; hSI; hPTI;  
 KW D2H; hPEPTI; human; GI tract receptor; sucrose-isomaltase complex;  
 KW intestinal peptide-associated transporter; hypertension; diabetes;  
 KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;  
 KW therapeutic agent delivery; therapy; probe.  
 XX Homo sapiens.  
 OS WO9851325-A2.  
 PN 19-NOV-1998.  
 XX 15-MAY-1998; 98WO-US10088.  
 PF 15-MAY-1997; 97US-0046595.  
 XX (CYTO-) CYTOGEN CORP.  
 PA (ELAN-) ELAN CORP PLC.  
 XX Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin IJ;  
 PI Omahony DJ, Patterson CA, Singleton J;  
 XX WPI; 1999-009568/01.  
 DR New proteins that bind specifically to receptors in the  
 XX gastro-intestinal tract and related nucleic acid - chimaeras and  
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 ; Search time 61.7736 Seconds  
(without alignments)  
3772.184 Million cell updates/sec

Title: US-10-043-418-3  
Perfect score: 4768  
Sequence: 1 MKKIPSLALGILLVFLLOY.....VAEISGLNLLGREFKLVLH 903

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL23.\*

1: sp\_archaea.\*

2: sp\_bacteria.\*

3: sp\_fungi.\*

4: sp\_human.\*

5: sp\_invertebrate.\*

6: sp\_mammal.\*

7: sp\_mhc.\*

8: sp\_organelle.\*

9: sp\_phase.\*

10: sp\_plant.\*

11: sp\_rodent.\*

12: sp\_virus.\*

13: sp\_vertebrate.\*

14: sp\_unclassified.\*

15: sp\_rvirus.\*

16: sp\_bacteriap.\*

17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2830.5	59.4	902	10 Q9LYF8	Q9LYF8 arabidopsis
2	2825.5	59.3	902	10 Q22444	Q22444 arabidopsis
3	2533.5	53.1	879	10 Q9LYI2	Q9LYI2 hordeum vul
4	2233	46.8	910	10 Q8VWV9	Q8VWV9 pinus pinas
5	2140.5	44.9	916	10 Q9AVC3	Q9AVC3 physcomitre
6	2129	44.7	907	10 Q9ZP26	Q9ZP26 arabidopsis
7	2129	44.7	915	10 Q9S7Y7	Q9S7Y7 arabidopsis
8	2083	43.7	928	10 Q9LGC6	Q9LGC6 oryza sativ
9	2046.5	42.9	928	10 Q9LEEC9	Q9LEEC9 solanum tub
10	2031	42.6	935	10 Q9ZP04	Q9ZP04 tropaeolum
11	1956	41.0	855	10 Q9LZT7	Q9LZT7 arabidopsis
12	1568.5	32.9	932	13 Q73626	Q73626 coturnix co
13	1535	32.2	953	11 Q8BG16	Q8BG16 mus musculu
14	1534	32.2	953	11 Q91Z45	Q91Z45 mus musculu
15	1501.5	31.5	952	4 Q81WE7	Q81WE7 homo sapien
16	1474	30.9	937	6 Q9MTM4	Q9MTM4 bos taurus

17	1445	30.3	992	3 Q9UV08	Q9UV08 emericeia
18	1444.5	30.3	995	3 Q9URX4	Q9URX4 schizosacch
19	1386	29.1	873	13 Q73632	Q73632 coturnix co
20	1371	28.8	1734	4 Q8TE24	Q8TE24 homo sapien
21	1355.5	28.4	955	3 Q9C1S7	Q9C1S7 emericeia
22	1186	24.9	1743	5 Q19004	Q19004 caenorhabdi
23	1160.5	24.3	955	5 Q21750	Q21750 caenorhabdi
24	1137.5	23.9	920	5 Q9NFY8	Q9NFY8 penaeus van
25	1041	21.8	856	5 Q20722	Q20722 caenorhabdi
26	881.5	18.5	921	10 Q9FN05	Q9FN05 arabidopsis
27	877	18.4	728	2 Q9RH22	Q9RH22 alicyclobac
28	849	17.8	914	4 Q81ZM5	Q81ZM5 homo sapien
29	848	17.8	925	4 Q8TET4	Q8TET4 homo sapien
30	846	17.7	914	4 Q81ZM4	Q81ZM4 homo sapien
31	845.5	17.7	746	16 Q8XIN9	Q8XIN9 clostridium
32	843.5	17.7	991	10 Q93Y12	Q93Y12 arabidopsis
33	842	17.7	751	16 Q81WZ0	Q81WZ0 thermoanaer
34	832	17.4	769	4 Q81WZ0	Q81WZ0 homo sapien
35	829.5	17.4	653	11 Q8BYM0	Q8BYM0 mus musculu
36	825	17.3	763	16 Q92F84	Q92F84 listeria in
37	820	17.2	779	16 Q8YN00	Q8YN00 anabaena sp
38	811.5	17.0	919	10 Q24375	Q24375 solanum tub
39	810.5	17.0	959	10 Q9LUG2	Q9LUG2 arabidopsis
40	805.5	16.9	944	4 Q14697	Q14697 homo sapien
41	805	16.9	966	11 Q08794	Q08794 mus musculu
42	804	16.9	966	4 Q9P0X0	Q9P0X0 homo sapien
43	803.5	16.9	944	11 Q8BHN3	Q8BHN3 mus musculu
44	800.5	16.8	763	16 Q8YAE7	Q8YAE7 listeria mo
45	799	16.8	944	6 P79403	P79403 sus scrofa

#### ALIGNMENTS

#### RESULT 1

ID	Q9LYF8	PRELIMINARY;	FRT;	902 AA.
AC	Q9LYF8;			
DT	01-OCT-2000 (TRENBLrel. 15, Created)			
DT	01-OCT-2000 (TRENBLrel. 15, Last sequence update)			
DT	01-MAR-2003 (TRENBLrel. 23, Last annotation update)			
DE	Alpha-glucosidase 1 (AT5G11720/T22P22_110).			
GN	T22P22_110.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsi.			
OX	NCBI_TaxID=3702;			
RN	[1]			
RA	SEQUENCE FROM N.A.			
RA	Bevan M., Hilbert H., Braun M., Holzer B., Brandt A., Duesterhoeft A.,			
RA	Bancroft I., Mewes H.W., Rudd S., Lemcke K., Mayer K.P.X.;			
RL	Submitted (APR-2000) to the ENBL/GenBank/DBJ databases.			
RN	[2]			
RA	SEQUENCE FROM N.A.			
RA	EU Arabidopsis sequencing project;			
RL	Submitted (APR-2000) to the ENBL/GenBank/DBJ databases.			
RN	[3]			
RA	SEQUENCE FROM N.A.			
RA	Shinn P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,			
RA	Banh J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D.,			
RA	Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,			
RA	Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,			
RA	Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,			
RA	Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,			
RA	Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,			
RA	Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;			
RT	*Arabidopsis cDNA clones*;			
RL	Submitted (AUG-2001) to the ENBL/GenBank/DBJ databases.			
RN	[4]			
RA	SEQUENCE FROM N.A.			
RA	Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,			
RA	Chan M.M., Chang C.H., Dale J.M., Hayashizaki Y., Hsuan V.W.,			



```
RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
RA Lee J.M., Lin J., Miranda M., Narueaka M., Nguyen W., Palm C.J.,
RA Quach H.L., Sakurai T., Satou M., Seki M., Southwick A., Talm C.C.,
RA Toriumi M., Wallender E.K., Wong C., Wu H.C., Yamada K., Yu G.,
RA Yuan S., Shinzaki K., Davis R.W., Theologis A., Ecker J.R.;
R "Arabidopsis ORF clones";
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL163814; CAB87690.1; -
DR EMBL; AY053414; AAK96644.1; -
DR EMBL; BT002222; AAN72333.1; -
DR InterPro; IPR000322; Glyco_hydro_31.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
SQ SEQUENCE 902 AA; 101118 MW; D56887DAA000B3B CRC64;

Query Match 59.4%; Score 2830.5; DB 10; Length 902;
Best Local Similarity 59.4%; Pred. No. 1.6e-205;
Matches 540; Conservative 138; Mismatches 205; Indels 23; Gaps 8;

QY 6 PSLALGILLVFLQYLQVLVAGISTSENDPEGVIGYKVKSVKVDGSTRSLTALPOLVQNS 65
DB 8 PNIFIVVVFSLR---SSQVLEEEESTVVGIVVRSVGVDN-NRQVLTAKLDLIRPS 63
QY 66 SVYGPDIQLLSITASLESNDRLRVRITDAKRRWEIPDNILHR---HQPPPPPHSLSSL 122
DB 64 SVYAPDIKSLNLHVSLETSERLRIRITDSSQORWEIPETVIPRAGNHSP-----RRFST 117
QY 123 YRTLLSSPTNRRKILLSHPNSDLTFSLINTTPGFTISRKSTHDVLFDATPDPTNPTF 182
DB 118 EEDGGSPENN---FLADPSSDLVFTLHNTTPGFVSRRSSGDLFDTPSPDSSEN 173
QY 183 LIFDQVHLTSSLPGRTRAHIYGLGHSKPTFQLAHQNTLTMRADIPSSNPVDNLYGSH 242
DB 174 FIFKQFQLQSALPENRSNLYGEGHTKRSFRIIPGETMTLWNADTGSNPVDNLYGSH 233
QY 243 PFYMDVRSS---PVAGSTHGVLNLSNGMDVEYTGNRITTKVIGIIDLFFAGPSPOV 299
DB 234 PFYMDVRSGKNEEAGTTHGVLLNLSNGMDVKYEGHRTYNNVIGVIDLYVFAGPSPEW 293
QY 300 VEQFTRVIGRPAPMPYAFGQCRQYGHVDYVTELQSVVAGYAKAKIPLVNMWTDIDYMDA 359
DB 294 MNQYTELIGRPAPMPYMSFGPHQCRYGYKVDLYVVDGYAKAGIPLVNMWTDIDYMDG 353
QY 360 YKDFTLDPNPLDMKKFVNHLKNGQKYYVLLDPGISTNKTYETIYIRGKMHDFLKN 419
DB 354 YKDFTLDPNPFEDKMQSFVDTLHKNQKYYVLLDPGIGVDSSYGTYNRGMEADVIFKRN 413
```

```
DB 774 SRLENISELFDGLENLWAGCGNDRDWTLVKFCYVTKSVVLRSEVVNPEYASOKW 833
QY 838 VIDKVTILGLRKGKINGYVTRTGAVTGRKDGSKLKSTPDRKG-EFTVAEISGLNLLGR 896
DB 834 SIGKTVFVGFEVNVKTYEVRTSERLRSRISLIKTVDNDPRFLSVESKLSLIVGK 893
QY 897 EFKLVL 902
DB 894 KFERL 899

RESULT 2
O22444 PRELIMINARY; PRT; 902 AA.
AC O22444;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Alpha-glucosidase 1 (EC 3.2.1.20).
GN AGLU1
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Landsberg;
RX MEDLINE=98045620; PubMed=9411456;
RA Monroe J.D., Hall B.D., Gough C.M., Stephen A.L.;
RT "Nucleotide sequence of an alpha-glucosidase gene (Accession No.
RT AF014806) from Arabidopsis thaliana (PCR97-141).";
RL Plant Physiol. 115:863-863(1997).
DR EMBL; AF014806; AAB82656.1; -
DR InterPro; IPR000322; Glyco_hydro_31.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 902 AA; 101106 MW; 71F2AB9F4D8D22EA CRC64;

Query Match 59.3%; Score 2825.5; DB 10; Length 902;
Best Local Similarity 59.3%; Pred. No. 3.8e-205;
Matches 539; Conservative 138; Mismatches 206; Indels 23; Gaps 8;

QY 6 PSLALGILLVFLQYLQVLVAGISTSENDPEGVIGYKVKSVKVDGSTRSLTALPOLVQNS 65
DB 8 PNIFIVVVFSLR---SSQVLEEEESTVVGIVVRSVGVDN-NRQVLTAKLDLIRPS 63
QY 66 SVYGPDIQLLSITASLESNDRLRVRITDAKRRWEIPDNILHR---HQPPPPPHSLSSL 122
DB 64 SVYAPDIKSLNLHVSLETSERLRIRITDSSQORWEIPETVIPRAGNHSP-----RRFST 117
QY 123 YRTLLSSPTNRRKILLSHPNSDLTFSLINTTPGFTISRKSTHDVLFDATPDPTNPTF 182
DB 118 EEDGGSPENN---FLADPSSDLVFTLHNTTPGFVSRRSSGDLFDTPSPDSSEN 173
QY 183 LIFDQVHLTSSLPGRTRAHIYGLGHSKPTFQLAHQNTLTMRADIPSSNPVDNLYGSH 242
DB 174 FIFKQFQLQSALPENRSNLYGEGHTKRSFRIIPGETMTLWNADTGSNPVDNLYGSH 233
QY 243 PFYMDVRSS---PVAGSTHGVLNLSNGMDVEYTGNRITTKVIGIIDLFFAGPSPOV 299
DB 234 PFYMDVRSGKNEEAGTTHGVLLNLSNGMDVKYEGHRTYNNVIGVIDLYVFAGPSPEW 293
QY 300 VEQFTRVIGRPAPMPYAFGQCRQYGHVDYVTELQSVVAGYAKAKIPLVNMWTDIDYMDA 359
DB 294 MNQYTELIGRPAPMPYMSFGPHQCRYGYKVDLYVVDGYAKAGIPLVNMWTDIDYMDG 353
QY 360 YKDFTLDPNPLDMKKFVNHLKNGQKYYVLLDPGISTNKTYETIYIRGKMHDFLKN 419
DB 354 YKDFTLDPNPFEDKMQSFVDTLHKNQKYYVLLDPGIGVDSSYGTYNRGMEADVIFKRN 413
```







GN XYL1 OR F24J5.20.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA Sampedro J., Sierro C., Villa T.G., Revilla G., Zarra I.;  
 RT "Cloning and expression pattern of an alpha-xylosidase gene from  
 RT Arabidopsis thaliana";  
 RL Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,  
 RA Lee J.M., Li J., Gonzalez A., Liu K., Vayenberg M., Sakano H.,  
 RA Chin C., Choi E., Chou J., Alcafi H., Araujo R., Brooks S.,  
 RA Buehler E., Chao Q., Conn L., Conway A.B., Dunn P., Hansen N.,  
 RA Howing B., Huizar L., Khan S., Kim C., Palm C., Rowley D., Shinn P.,  
 RA Walker M., Davis R.W., Ecker J.R., Federpiehl N.A., Theologis A.;  
 RT "The sequence of BAC F24J5 from Arabidopsis thaliana chromosome 1";  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,  
 RA Bower L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,  
 RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,  
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,  
 RA Nguyen M., Onodera C.S., Palm C.J., Pham C.J., Quach H.L., Sakurai T.,  
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,  
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,  
 RA Ecker J.R.;  
 RT "Arabidopsis cDNA clones";  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Kim C.J., Chen H., Cheuk R., Shinn P., Bowser L., Carninci P.,  
 RA Chan M.M., Chang C.H., Dale J.M., Hayashizaki Y., Hsuan V.W.,  
 RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,  
 RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,  
 RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,  
 RA Tang C.C., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G., Yuan S.,  
 RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;  
 RT "Arabidopsis ORF clones";  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF144078; AAD37363.1; -  
 DR EMBL; AC008075; AAD49987.1; -  
 DR EMBL; AY057482; AAL09716.1; -  
 DR EMBL; BT002675; AAO11591.1; -  
 DR InterPro: IPR000322; Glyco\_hydro\_31.  
 DR Pfam: PF01055; Glyco\_hydro\_31; 1.  
 DR PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
 KW Signal.  
 FT SIGNAL.  
 FT CHAIN 1 27 POTENTIAL.  
 FT CHAIN 124 915 ALPHA-XYLOSIDASE.  
 SQ SEQUENCE 915 AA; 102398 MW; 28F9610D8D7EA657 CRC64;  
 Query Match 44.7%; Score 2129; DB 10; Length 915;  
 Best Local Similarity 44.9%; Pred. No. 2.5e-152;  
 Matches 417; Conservative 167; Mismatches 282; Indels 62; Gaps 12;  
 QY 7 SLALGILLVFLQLVAGISTENDEPGVIGYGVKSVKVDGTRRSITALPQLVKSS 66  
 Db 9 AFSLSULLALICF-----SPTOSYKTIHGKYL--VSIEESDGGFVGLVQKQNK 59  
 QY 67 VYGPDTQLLSITASLESNDRLAVRTDAKRRWEIPDNILRHQPPPPPHSLSLYRTL 126  
 Db 60 IYGSODITLRLVKGHTDSRLAVHITDAKQWEPYNLLPREQP-----QVGKVKSS 114  
 QY 127 LSPPTNRKILLSHPNSLTSLINTPTPGFTISRKSTHDLVFOATPDPTNPTFLIFI 186

Db 115 RKSPIT-----VQFISGSELIFS-YTTDPTFAVKRSNHELTFTNT-----SSLVFK 161  
 QY 187 DQYLHLTSSLPGTRAHYIYGLGHSKPT-FQLAHNQTLTWRAADIPSSNDVLYGHPFY 245  
 Db 162 DQYLEISTSLP-KEASLYGLGNSQANGIKLVENEPYTLTYEDVSAINLNTLDYLGHPMY 220  
 QY 246 MDVRSVPVAGSTHGVLLNSNGMDVEYTGRIYKVIIGGIIDLYFFAGSPGQVVSQFTR 305  
 Db 221 MDNRVGGKAYAHAVULLNSNGMDVYRGDSLYIKVIGGVDFYFIAGSPPLNVDDYIQ 280  
 QY 306 VIGRPAPMPYAFGQOCRYGYHDVYELQSVVAGYAKAKIPILEVMTDIDYMDAYDFTL 365  
 Db 281 LIGRPAPMPYSLGPHQCRWGHNSLVVEDVDNYYKAKIPLDVIWDDHMDHKDFTL 340  
 QY 366 DPVNFPLDKKKFVNHLHNGQKVVILDPGISTNTKTYETIRGMKHDFLKRNGKPYLG 425  
 Db 341 NPVAYPRAKLLAFLDKIKHIGMKYIIVNDPGIGVNASYGTQFRAAADVFIKYEKPFLLA 400  
 QY 426 SVMFGPVYFPDFLKPSALTFTWDEIKRFLNLLPVDGLWIDMNEISNFS----- 474  
 Db 401 QVWPGPVYFPDFLNPKTVSWGDEIKRFDLVIDGLMIDMNEVSNPCSLGLCTIPEGKQC 460  
 QY 475 -SPPIPG-----STLDNPPYKINNSGVMLPIINKTIPPTAMHYGDIPEYVNH 521  
 Db 461 PSGEFGWVCCLDCKNITKTRWDDPPYKINATGVAPVGFKTIATSATHYNGVREYDAHS 520  
 QY 522 LFGYLEARVTRALILKTEKRPVLSRSTFSSGSKYTAHTGDNAATWDLVYSISMLD 581  
 Db 521 IYGFSETIATHKGLLVQGGKRPILSRSTFVSGVQYAAHWTDGNOGTWQSLQVSIITMLN 580  
 QY 582 FGLFGIPWVGADICGFLNGTTEELCRWLTOLGAFYPPFSRDHSLGTTYOELVWESVAAS 641  
 Db 581 FGIFGVPWSDICGFPYPPTEELCRWIEVGAFYPPSRDHANYSPRQELYQWDIVADS 640  
 QY 642 ARKVLGLRYTLPLFYTLMYEALNGIPIARPLFFSPDDIKTYGTSOFLKGGWVWSP 701  
 Db 641 ARNALGMRKILPLFLTYLTYEAHMTGAPIARPLFFSFPEYTECYGNSRQFLLGSSPMISP 700  
 QY 702 VLKPGVSVTVAPRGNWPDLDYTSVASTAGRYVTLTSLAPDHINHVHIOGNILAMQOK 761  
 Db 701 VLSQKTEVEALPPGSHWHFMFDMTQAVSKNGKRVTLPAFLNVNHLVQNTILTQQG 760  
 QY 762 AMTTOARKTFFHLVVM-----SDCGASFGELFLDDGVVEVTMGVNRGKTFVKFIAASAKQ 818  
 Db 761 GLISKDAITTFSLVIAFPAGASEGVATGKYLDEDELPEMKLNGQSTVVDVYASVNG 820  
 QY 819 TCIITSDVSGEFAVSKQWIDKVTILGLKGTNGYVTRTGAVTRKGDKSKLKST--- 875  
 Db 821 TKXWMSQVKEGKFAISKGWIEKSVLGLRGAGQVSEIQINGSPMTKIEVSSKEHTYVI 880  
 QY 876 ---PDRKGEFIVAEISGLNLLIGREFKL 900  
 Db 881 GLEDEENKSVWVEVRGLEMLVGKDFNM 908  
 RESULT 8  
 Q9LGC6 PRELIMINARY; PRT; 929 AA.  
 AC Q9LGC6;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE EST AU092739 (CS3221) corresponds to a region of the predicted  
 DE gene.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;

RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC  
clone: P0504H10."  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AP002526; BAA39366.1; -.  
DR Gramene; Q9LGC6; -.  
DR InterPro; IPR003322; Glyco\_hydro\_31.  
DR Pfam; PF01055; Glyco\_hydro\_31; 1.  
DR PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
SQ SEQUENCE 929 AA; 102550 MW; 536ELD0B9D7P97BF CRC64;

Query Match 43.7%; Score 2083; DB 10; Length 929;  
Best Local Similarity 46.7%; Pred. No. 8e-149;  
Matches 423; Conservative 150; Mismatches 269; Indels 64; Gaps 15;  
QY 36 IGYGVKSVKVDSTGTRSLTALPOLVKNSSVGPDIQLLSITASLESNDRLVRITDAK 95  
DB VSGGKVLVSL-VEHEGALVGYLOVKRTSTYGPDIPLRLYVKHETKDIRVQITDAD 97  
QY 96 HRRMEIPDNILHRHOPPPPHSLSSLYRTLLSSFTTNRKKILLSHNSDLTFLSLNTP 155  
DB KPRMEVPYLLQR-EPAPP-----VTGGRITGVPPFAAGEYPGCELVFT-YGRDP 144  
QY 156 FGFTISRSKTHDVLDPATDPDTPNPNFTLIFIDQYHLTSSLPGRTHIYGLGHSKP-TF 214  
DB FWFVHRKSSREALPNTSCGA-----LVFKDQVIEASTSLPRDAA-LYLGENTQPGGI 197  
QY 215 CLAHKQTLTMAADIPSSNPVNLVGSHPFMYDVRSSPVAGSTHGVLILNSGMDVEYTG 274  
DB RLRPNDPVTIYTTDISALNLTDLVGSHPVYVUDRSRGCHGVAAVLLNSGMDVYRG 257  
QY 275 NRITYKVIIGGIDLYFFAGSPGVQVEQTRVIGRPAPMPYKAFQOCRYGYHDVYELQ 334  
DB SLTYKVIIGLLDFYLFSGPTPLAVDQYTSWIGRPAPMPYKAFQOCRYGYHDVYELQ 317  
QY 335 SVVAGYAKAPLEVMWTDIDYMDAYKFTLDVNPFLDKMKFFVNNLHKGQKVVLLD 394  
DB GVVEYRQAQPLDVIWDDHMDAOKFTLDVNPYRPKLLFLELDKTHAQCKMYIILID 377  
QY 395 PGISFNKYETVIRGMKHDFVLKRGKPYLGSVMPGVYFPELKPSTALTFTWDEIKRFL 454  
DB PGAVNNTYGVYQRMQGDVFIKDGKPYLAQVMPGVYFPELNPNGVSWWIDEVRRFH 437  
QY 455 NLLPVDGLWIDMNEISFISSP-----PIPGSTL-----DNPPYK 489  
DB DLVPVDGLWIDMNEASNETCKEIPTHLCPLPNTTTPWVCCLDCKNLNTRWDEPYK 497  
QY 490 INNSGVMLPIINKTPTTAMHYGDIPEYNNVHNFYGLYEARVTRAAIKLTKRPFVLSRS 549  
DB INASQTLARLGNTIATSTATHNGILEYNAHSLYGSQAIAHQALOGKQKRPFILTRS 557  
QY 550 TFGSGKYTAHWTGONATWDLVYSIPSLMDPLFGIPMPMGADICGFLGNTTEELCRW 609  
DB TVGSGAVAAHWTGONKGTWELRYSISTMLNFGIFGMPMGADICGYPQTEELCRW 617  
QY 610 IOLGAFYFSDRHSSLGTTYQELRWESVAASARKVLGLRYTLFPYLYTLMYEALNGIP 669  
DB IELGAFYFSDRHANFASPRQELVYVESVAKSARNALGMRYELLFYLTYLTAQHLTGA 677  
QY 670 IARPLFFSPDDIKTYGSSQFLKQGVMSVPLKGVVSVTAYPPRGWFDLDPYTRSV 729  
DB VARPVFFSPDPPTCYGLSTQYLLGASVMSVPLEGATSVSAMPPGGSWYNLEFDTTKV 737  
QY 730 TASTCRYVTLSPAPPHINVIHIOEGNILAMOGKAMTTOAARKTPFHLLVMSDCGA---S 785  
DB VSRGEGAVKLDAPLNEINHVHVFQNTILPMQRGGTISKERATPP-TLVVAPFGATEABA 796  
QY 786 FGELELDGGEVMTGVNRRKMTFFVAFIAASAKQTCIITSDVVSGBFASQKWVIDKVTIL 845  
DB ECAVYVDDDEREMVLAECQATYVRYFATVRCAVTVRSEVELGYSYQKGLLIEKLSVL 856  
QY 846 GLRKGK-----KINGYVTRQAVTR---KDGKSKLSTPDRKG---EFIVAEISGINULL 894  
DB GL-EGTGRDLAVHVDGANATAIATSRPVFAGAEALHGHHRDVEGHKXSMVVEVGLAFL 915

QY 895 GREPKL 900  
DB 916 GKSFTW 921

## RESULT 9

Q9LEB9 PRELIMINARY; PRT; 928 AA.  
AC Q9LEB9  
DT 01-OCT-2000 (TREMREL. 15, Created)  
DT 01-OCT-2000 (TREMREL. 15, Last sequence update)  
DT 01-MAR-2002 (TREMREL. 20, Last annotation update)  
DE Alpha-glucosidase (EC 3.2.1.20).  
GN MAL2.  
OS Solanum tuberosum subsp. tuberosum.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; lamiids; Solanales; Solanaceae; Solanum.  
OX NCBI\_TaxID=90692;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Record; TISSUE=Tuberising stolon tip;  
RX MEDLINE=21362238; PubMed=11469591;  
RA Taylor M.A., Ross H.A., McRae D., Wright F., Viola R., Davies H.V.;  
RT "Copy-DNA cloning and characterisation of a potato alpha-glucosidase:  
expression in Escherichia coli and effects of down-regulation in  
transgenic potato";  
RL Planta 213:258-264 (2001).  
DR EMBL; AJ27244; CAB96077.1; -.  
DR InterPro; IPR003322; Glyco\_hydro\_31.  
DR Pfam; PF01055; Glyco\_hydro\_31; 1.  
DR PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
KW Glycosidase; Hydrolase.  
SQ SEQUENCE 928 AA; 104687 MW; CEAC8634F6ED1820 CRC64;

Query Match 42.9%; Score 2046.5; DB 10; Length 928;  
Best Local Similarity 44.6%; Pred. No. 4.7e-146;  
Matches 415; Conservative 151; Mismatches 264; Indels 101; Gaps 16;

QY 25 ISTSENDPEGVIGYGVKSVKVDSTGTRSLTALPOLVKNSSVGPDIQLLSITASLESN 84  
DB IAIEESPDDGLIGY-LKVK-----KKNYIGPDIPLNQLYVYKHETD 77  
QY 85 DRLRVRITDAKRWEIPDNILHRHOPPPPHSLSSLYRTLLSSFTTNRK---ILLS- 140  
DB NLRIRHITDAKORWEVPYLLPRESP-----SLKQTKGS-----RKQFSLLSN 124  
QY 141 --HPNSDLTSLINTTTPFGFTISRSKTHDVLDPATDPDTPNPNFTLIFIDQYHLTSLPG 198  
DB QEYSGNELMFS-YTSDPFSSVKRSKNGQTLFNSSESDSDPYSLNLFKQYLEISTKL- 182  
QY 199 TRAHYGLGHSKP-TFOLAHNQTLMRAADIPSSNPVNLVGSHPFMYDVRSSPVAGST 257  
DB KDAASLYLGLGENTOPHGKIYPNDPYTLTYTDDQSSINLMDLYGSHPMYMDLRNVGEAYA 242  
QY 258 HGVLLNLSNGMDVEYTNRYTKVIGIIDLFFAGSPGVQVEQTRVIGRPAPMPYMA 317  
DB HAVLLNLSNGMDVEYTNRYTKVIGIIDLFFAGSPGVQVEQTRVIGRPAPMPYMA 302  
QY 318 FGFQCRKYGYHDVYELGSSVAGYAKAKIPLVNMWTDIDYMDAYKFTLDVNPFLDRMKK 377  
DB FGFQCRKWYVHNLVSVIEDVIANKKAKIPLDIVWDDHMDGKKDFTLHPINTYFGPKLRA 362  
QY 378 FVNHLNKGQYVILDPGISTTKTYIRGKMKHDFVLKBNCKPYLGSVMPGVYFPDF 437  
DB FLEKIHAEAGHYIVINDPGIGVKNKSYGTQYRGANDVFIKYEKGFPLAQWPGAVHFPDF 422  
QY 438 LKPSALTFTWDEIKRFLNLLPVDGLWIDMNEISNF-----I 473  
DB LNPVTWEGDELRPHELAPIDGLWIDMNEVSNFNCGLCTIPEGRICPNGTGPGWICCL 482  
QY 474 SSPPIPGSTLNDPPYKINNSGVMLPIINKTPTTAMHYGDIPEYNNVHNFYGLYEARVTR 533







```

Db 326 LTKRTIGVDFYFLGPDPMNVVQQOEVIGFPAMPPLNALGHLCRWGYSNETWQT 385
Qy 337 VAGYAKAKIPILEVWMTDIDYDAYKDFLDPVNFPLDKMKKFNVLNHNKGQKYYVILDPG 396
Db 386 AKAMRNFIQPDQAQWNOIDYMDGYRDETFDPQKFA--SLPSLVEDLKHGHYVILDPG 443
Qy 397 ISTNK---TYETIRGMKHDFL-KNGKPYLGSVMPGVVFPDFLAKPSALTPTWDEIKR 452
Db 444 ISSSTPRGSYWPFDGLRGFLNLTQQTLIGQWPGYTAYPDPSTNDTHQWLENLQR 503
Qy 453 FLNLLPVDGLWIDMNEISNFI--SSPIPGSTLNNPPYK---INNSGYMLPIINKTIPPT 507
Db 504 FHTHVPFDGLAIDMNEPSNFDNFBEGCPGELSDSPYTPAVLGN-----LTKATVCAS 558
Qy 508 AMHVGDIPEYNVHNLFGYLEARVTRAAULIKUTEKRPFLVLSRSTFGSGSKYTAHWGDNAA 567
Db 559 AEQNASV-HYNLHNLGLKEAEATASALIRIGKRPFVISRSTFSPQGRYSGHWLGNRS 617
Qy 568 TWNLDVYSIPMLDPLFGIIPMGADICGFLGNTTEELCREWIOLGAFYPSRDHSSLGT 627
Db 618 QWKMYNIPGMLSESLGIFLVGADICGFSGSSEELCTRWMLQAGAFYPSRHHQNE 677
Qy 628 TYQELRWESVAASARK-VLGLRYTLFPYTYLMEYEAQLNGIPIARPLFFSFPDDIKTYG 686
Db 678 KAQDPTAFSPARTAMKDALTRYSLFLYTLPHRAHLQGETVARPLFFFPMDVATYG 737
Qy 687 ISSQELLGKGMVSVLKGVSUVTAYPRGNWFDLPDTRSVTASTGRVYTLSPADHI 746
Db 738 LDRQFLWGQSLVTPVLPFGADSVLGYPPQGVYDF--YTGSVNSSGEMLKLSAPLDHL 795
Qy 747 NVHIQEGNILAMQKAMTQAAKTPPHLLVMSDCGASFGELFLDDGVEVTMGVNRGW 806
Db 796 NLHREGSILTPKQIITSKATRNPLHLIVALSTRATAMGDLFWDDG-ESLDTFEQNY 854
Qy 807 TFKVPIAASQTCIITSDDVSGEFAVSQKWIWDKVTILGKRGTKINGYTVRTGAVTRK 866
Db 855 SYLVFNATEN---IFTSNVLHASTEADTV-TIDAVSFYQVE-----PPSKVLLD 900
Qy 867 GDKSKLKSTPDRKGEFIVAEISGLNLLLGREFKL 900
Db 901 GQEKPFSLYDQ-----VLTVSLGLVLVSQGFSL 929

RESULT 13
Q8BG16 PRELIMINARY; PRT; 953 AA.
AC Q8BG16;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DE Glucosidase.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN NCBI_TaxID=10090;
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J, and NOD; TISSUE=Heart, and Thymus;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RT "Analysis of the mouse transcriptome based on functional annotation of
RL Nature 420:563-573 (2002).";
DR EMBL; AK052211; BAC34886.1; -.
DR EMBL; AK088481; BAC40382.1; -.
SQ SEQUENCE 953 AA; 106247 MW; 956B89685FB5FF81 CRC64;

Query Match 32.2%; Score 1535; DB 11; Length 953;
Best Local Similarity 38.4%; Pred. No. 2.9e-107;
Matches 341; Conservative 136; Mismatches 312; Indels 99; Gaps 23;

```

```

Qy 40 YKVKSVKVDSTRRSLTALPOLVKNSVYGP-DIQLLSITASLESNDRLRVRITDAKRR 98
Db 136 YLENL---SSTESGYTA--TLTRTPSPFPKDVLTQLQVLEVMETDSRLHFKIKDPASR 190
Qy 99 WEIPDNILHRHOPPPPHSHSLRYTLSSPTTNRKILLSHPNSDLTFTSLINTTTPFG 158
Db 191 YEVP-----LETPR-----VLQAPSPLYSVFSEEPGV 220
Qy 159 TISRSKTHDVLFDATPDPTNPNTFLIFIDYHLHLSLPGTBAHIYGLGESHKPTFQLAH 218
Db 221 IVRKLGRVLLNTTVP-----LFFADQFLQLSTLPS--OHITGLGEHLSLMLSTD 272
Qy 219 NOTLTMRAADIPSSNPVNLGSHPHFYMDVRSSPVAGSTHGVLLNSNGMDVEYTCN-RI 277
Db 273 WARIILNRDTPSQ-GTNLYGSHFYLALD--GGLAHGVFLNNSAMVDILQSPAL 328
Qy 278 TYKVGIIIDLYFFAGSPGQVVEQFTRVIGRAPMPYWAFOQCRYGVHDVYELQSV 337
Db 329 TWRSTGGIIDVYVFLGPEPKSVVQVLDVVGYPFPMPYMGFLGFLCRWGYSTAIVRQV 388
Qy 338 AGYAKAKIPILEVWMTDIDYDAYKDFLDPVNFPLDKMKKFNVLNHNKGQKYYVILDPG 397
Db 389 ENMTRTHFPLDVQMDLDYMDARRDFTFNQDSFA--DPPDMVRELHQDGRRRYMMIVDPA 446
Qy 398 ST---NKTETVIRGMKHDFL-KENGKPYLGSVMPGVVFPDFLKPASALTPTWDEIKR 453
Db 447 SSAGPAGSVRPYDEGLRGVFTNETQPLICKWPGTTAFDPDTPNETLDMQWVSEF 506
Qy 454 LNLPLVDGLWIDMNEISNFI--SSPIPGSTLNNPPYKINNSGVMLPIINKTIPPTAMH 511
Db 507 HAQVPDPGMWLDNNEPSNFRVSGSQGCPNNELENPPV---FGVVGILQAAATCASSHQ 563
Qy 512 GDPIEYVNLFCYLEARVTRAAULIKUTEKRPFLVLSRSTFGSGKYTAHWGDNAAWTD 571
Db 564 FLSTHNLHNLGLTEAIASSRALVKTGRTPVLSRSTFGHGRYAGHWTDGVRSSWEH 623
Qy 572 LVYSIPEMLDFGLFGIPMVGADICGFLGNTTBEELCRWIOLGAFYPSRDHSSLGTTYOE 631
Db 624 LAYSVDIILQNLGLVPLVGADICGPIGDTSEELCVRTQLGAFYPMFNHNDLNSVPE 683
Qy 632 LTRWESVAASA-RKVLGLRYTLFPYTYLMEYEAQLNGIPIARPLFFSFPDDIKTYGISQ 690
Db 684 PYRFSETAQQAARKAPALRYALLPYTLFHRHVRGDTVARPLEFPEDEPSTWSVDQ 743
Qy 691 FLGKGMVSVPLKGVSVVTAFFPRGNWFDL-----EDYTRSVTASTG 734
Db 744 LMGPAALLITPVLPEKTEVTGYFGKTYNQMVSVDSLGLTLPSPSSASSFRSAVQSKG 803
Qy 735 RYVTLSDPPDHINVHIQEGNILAMQKAMTQAAKTPPHLLVMSDCGASFGELFLDDG 794
Db 804 QWLTLEAPLDTINVHLREGYIIPLOQPSLTTTESRKQPMALAVALTASGEADGELFWDDG 863
Qy 795 VEVTMGVNRKWTFFKFIASAQTCIITSDDV-VSGEFAVSQKWIWDKVTILGKRGTKI 853
Db 864 ESLAV-LERGAYTLVTF---SAKNNTVINKLVKRVTKEGAEQ---LREVTVLGVA----- 911
Qy 854 NGYTVRTGAVTRKGDGSKLKSTPDRKGEFIVAEISGLNLLLGREFKL 900
Db 912 ---TAPQVLVSLNGIPVSNFTYSPDNKSLAI-----PVSLLMGELFQI 950

RESULT 14
Q91245 PRELIMINARY; PRT; 953 AA.
AC Q91245;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Similar to glucosidase, alpha, acid.
GN GAA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC010210; AAH10210.1; -.
DR MGD; MGI:95609; Gaa.
DR InterPro; IPR00322; Glyco_hydro_31.
DR InterPro; IPR000519; P_trefoil.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR Pfam; PF00088; trefoil; 1.
DR SMART; SM00018; PD; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
DR PROSITE; PS00025; P_TREFOIL; 1.
SQ SEQUENCE 953 AA; 106157 MW; 320B6D7354ASFF91 CRC64;

Query Match 32.2%; Score 1534; DB 11; Length 953;
Best Local Similarity 38.4%; Pred. No. 3.5e-107;
Matches 341; Conservative 135; Mismatches 313; Indels 98; Gaps 23;

QY 40 YKVKSVKVDGTRRSALTALPOLVKNSSVVGPP-DIQLLSITASLESNDRLRVITDAKRR 98
Db [1]
QY 136 YALENL-----SSTESGYTA--TLTRTPTFFPKDVLTLQLEVLMEIDSRHLKIKDPASGR 190
Db [1]
QY 99 WEIPDNILHRHQPPPPHSLSLYRTLSSPTTNRKILLSHPNNSDLTFSLINTTPEGF 158
Db [1]
QY 191 YEVP-----LETPR-----VLSQAPSPLYSEFSEEPGV 220
Db [1]
QY 159 TISRKSTHVDLFDATPDPTNPNTFLIFDIQYHLTSSLPCTRAHIYGLGHSKPTFQLAH 218
Db [1]
QY 221 IVRRLDGRVLLNTTVP-----LFFAQDLQLSTLSPS--QHITGLGEHLSPMLSTD 272
Db [1]
QY 219 NQTLTWRADIPSSNDVNLGSHHPYMDVRSSPVAGSTHGVLLNSNGMDVEYTG--RI 277
Db [1]
QY 273 WARIILWRDTPSQ--GTNLYGSHFPYLALED---GGLAHGVFLNSNAMDVILQSPAL 328
Db [1]
QY 278 TYKVIIGGIDLYPFAGSPGVVEQFTRVIGRPAPMPYMAFGFQOCRYGHDVYELQSV 337
Db [1]
QY 329 TWKSTGILDVYFLGPEKSVVQVLDVVGYPFPPYKGLGFLCRWGYSSTAIVRQVV 388
Db [1]
QY 338 AGYAKAKIPLVNMWDIDYMDAYKDTLPDVPFLDKMKKFFNNLHKNQKQYVILDDG; 397
Db [1]
QY 369 ENMTREHFDLQWMDLDYMDARRDTFNQDSFA--DFPDVYRELHQGRYMMIVDPA; 446
Db [1]
QY 398 ST---NKTYETIRGMKHDVFL--KENGKPYLGSMVPGPVYFDELKPSALTFTWDEIKRF 453
Db [1]
QY 447 SSAGPAGSYRPPDEGLRGVFTNETGQLKGVWFGTATFPDFTNPETLQWQDMSEF 506
Db [1]
QY 454 LNLPLVDGLWIDMNEISNFI--SSPPIPGSTLDNPPYKINNNGVMLPIINKTIPTTAMFY 511
Db [1]
QY 507 HAQVPFDGWLMDNNEPSNVRSGQCCPNNELENPPYV---PGVVGILQAATTCASSHQ 563
Db [1]
QY 512 GDIPENVNLGYLEARVTRAALIKLTKERPPVLSRSTFSGSKYTAHWTDNNAATND 571
Db [1]
QY 564 FLSTHYNLNLHLYGLTEAASSRALVKTRGTREPVISRSTFSGHGYAGHWGTGDRSSWEH 623
Db [1]
QY 572 LVYSIPSMDFGLFGIPMGADICGFLGNTTEELCRRMTQLGAFYPPFDRHSSLSGTTTQE 631
Db [1]
QY 624 LAYSPDILQFLLGVPLVGADICGFGIDTSELVCRVLTQLGAFYPPFWRHNDLNSVQOE 683
Db [1]
QY 632 LYRWESVAASA--RKVLGLRYTLPLYFYLMYBAQLNGIPIARLPFPFDPDIKTISSQ 690
Db [1]
QY 684 PYRFSETAQOAMKAFALRYALLPLYTLFHRARVGDVVARPLEFPEPDPSTWSVRQ 743
Db [1]
QY 691 FLILGKVMVSPVLKGVSVTAYFRGNWFDL-----FDYTRSVTAGT 734
Db [1]
QY 744 LWMGALLITPYLEPKTEVTGYFPKGTMYNNQVSVDSLGLTLPSPSASSFRSAVQSKG 803
Db [1]
QY 735 RYVTLAPPDHNHVIQENILAMQGMKMTTCARAKTPHLLVVMSCAGSGELFLDDG 794
Db [1]
QY 804 QWLTLEAPLDTINVHLREGYIIIPLOQPSLTSTTESRKQPMALAVALTASGEADGELFWDDG 863
Db [1]
```

```
QY 795 VEVTMGVNRGKWTFFVKFIAASAKOTCIITSDV-VSGEFAVSQKVIDKVTILGLRKGTKI 853
Db [1]
QY 864 ESLAV--LERGAYTLVTF---SAKNNTIVNKLVRVYKEGAELO---LREVTVLGVA----- 911
Db [1]
QY 854 NGYVTRTGAVTRKGDKSILKSTPDRKGEFIVABISGLNLLGLREFKL 900
Db [1]
QY 912 ---TAPTQVLSNGIPVSNFTYSPDNKSLAI-----PVSLMGLGELFQI 950
Db [1]

RESULT 15
Q8IWE7 PRELIMINARY; PRT; 952 AA.
AC Q8IWE7;
DT 01-MAR-2003 (TriEMBLrel. 23, Created)
DT 01-MAR-2003 (TriEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TriEMBLrel. 23, Last annotation update)
DE Glucosidase, alpha, acid (Pompe disease, glycogen storage disease type 11).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RC TISSUE=DuoDenum;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC040431; AAH040431.1; -.
SQ SEQUENCE 952 AA; 105323 MW; 6E2717BF201F469 CRC64;

Query Match 31.5%; Score 1501.5; DB 4; Length 952;
Best Local Similarity 38.0%; Pred. No. 1e-104;
Matches 341; Conservative 138; Mismatches 295; Indels 123; Gaps 26;

QY 40 YKVKSVKVDGTRRSALTALPOLVKNSSVVGPP-DIQLLSITASLESNDRLRVITDAKRR 98
Db [1]
QY 136 YALENL-----SSEMGYTA--TLTRTPTFFPKDILTLRDVMMETENRHLFTIKDPANRR 190
Db [1]
QY 99 WEIPDNILHRHQPPPPHSLSLYRTLSSPTTNRKILLSHPNNSDLTFSLINTTPEGF 158
Db [1]
QY 191 YEVPLETPHVSRAPSLYSVE-----FSEEPFGV 220
Db [1]
QY 159 TISRKSTHVDLFDATPDPTNPNTFLIFDIQYHLTSSLPCTRAHIYGLGHSKPTFQLAH 218
Db [1]
QY 221 IVRRLDGRVLLNTTVP-----LFFAQDLQLSTLSPS--QYITGLAEHLSPMLSTS 272
Db [1]
QY 219 NQTLTWRADIPSSNDVNLGSHHPYMDVRSSPVAGSTHGVLLNSNGMDVEYTG--RI 277
Db [1]
QY 273 WTRITLWNRDL--APTGANLYGSHFPYLALED---GGAHGVFLNSNAMDVVLQSPAL 328
Db [1]
QY 278 TYKVIIGIIDLXPFAGSPGVVEQFTRVIGRPAPMPYMAFGFQOCRYGHDVYELQSV 337
Db [1]
QY 329 SWRSTGILDVYFLGPEKSVVQVLDVVGYPFPPYKGLGFLCRWGYSSTAITRQVV 388
Db [1]
QY 338 AGYAKAKIPLVNMWDIDYMDAYKDTLPDVPFLDKMKKFP---VNNLHKNQKQYVILDD 394
Db [1]
QY 389 ENMTREHFDLQWMDLDYMDARRDTFNQDSFA--DFPDVYRELHQGRYMMIVD 443
Db [1]
QY 395 PGIISTN---KTYETIRGMKHDVFL--KENGKPYLGSMVPGPVYFDELKPSALTFTWDEI 450
Db [1]
QY 444 PATSSSGPAGSYRPPDEGLRGVFTNETGQPLKGVWFGTATFPDFTNPALAMWEDWY 503
Db [1]
QY 451 KRFLNLLPVDGLWIDMNEISNFI--SSPPIPGSTLDNPPYKINNNGVMLPIINKTIPTTA 508
Db [1]
QY 504 AEHFDQVFPDGMWIDNNEPSNFIKRGSDGCPNNELENPPYV---PGVVGTLQAATTCAS 560
Db [1]
QY 509 MHYGDIPENVNLGYLEARVTRAALIKLTKERPPVLSRSTFSGSKYTAHWTDNNAAT 568
Db [1]
QY 561 SHQFLSTHYNLNLHLYGLTEAASHRALVKARGTREPVISRSTFSGHGRVAGHWTDGVMSS 620
Db [1]
QY 569 WNDLVYSIPSMDFGLFGIPMGADICGFLGNTTEELCRRMTQLGAFYPPFDRHSSLSGTT 628
Db [1]
QY 621 WEQLASSVEIILQFLLGVPLVGADVCGFLGNTTEELCVRVLTQLGAFYPPFMRHNSLLSL 680
Db [1]
```

```
QY 629 YQELYRWESVAASA-RKVLGLRYTLPPYPTLMYEAQLNGIPARPLFFSFPDDIKTYGI 687
Db 681 PQEYPYSFSEPAQAMRKALTLRYALLPHLYTLPHQAHVAGETVARPUFLEFPKDSSTWTV 740
QY 688 SSQFLLGKGVVSPVLKPGVSVTAFFPRGNWFOLFVTRSVTA----- 731
Db 741 DHQLLWGEALLITPVLOAGKAETGCFPLGTWYDL--QTVPEALGSLPPPPAAPREPAZ 798
QY 732 -STGRYVTLSPDHHINVOEGNILAMOGKAMTQAAKTFPHLLVVMSCGASFGELF 790
Db 799 HSEGQWVTLPAPLDTINVHLRAGYIIPLOGPLTTESRQQPMALAVALTKEGEARGELF 858
QY 791 LDDGVEVTMGVNRGKWTFKFIAASAKQTCIITSDV-VSGEFVSVOKWVIDKVTILG--- 846
Db 859 WDDGESLEV-LERGAYTQVIFL---ARNNTIVNELVRVISEGAGLQ---LOKVTVLGVAT 911
QY 847 -----LRKGTKINGVTVRTGAVTRKGDKSKLKSTPDRKGEFIVAEISGLNLLGREF 898
Db 912 APQOVLSNGVPVSNFTY-----SPDTK-----VLDIC-VSLLMGEQF 947
```

Search completed: October 27, 2003, 10:31:02  
Job time : 66.7736 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 / Search time 12.5556 Seconds  
(without alignments)  
3382.161 Million cell updates/sec

Title: US-10-043-418-3

Perfect score: 4768

Sequence: 1 MKXKIPSLALGILLVFLQY.....VAEISGLNLLGREFKVLVLH 903

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4768	100.0	903	1	AGLU_SPTOL
2	3243.5	68.0	913	1	AGLU_BETVU
3	2422.5	50.8	877	1	AGLU_HORVU
4	1499	31.4	864	1	AGLU_MUGUA
5	1498	31.4	953	1	LYAG_MOUSE
6	1492.5	31.3	952	1	LYAG_HUMAN
7	1464.5	30.7	969	1	AGLU_SCHPO
8	1431.5	30.0	946	1	AMYG_CANAL
9	1410.5	29.6	985	1	AGLU_ASPOR
10	1401.5	29.4	985	1	AGLU_ASPNG
11	1379	28.9	993	1	YAJI_SCHPO
12	1374.5	28.8	1856	1	MCA_HUMAN
13	1374	28.8	958	1	AMYG_DEBOC
14	1354.5	28.4	1826	1	SUIS_RABIT
15	1323	27.7	1840	1	SUIS_RAT
16	1318.5	27.7	1826	1	SUIS_HUMAN
17	1315.5	27.6	1812	1	SUIS_SUNMU
18	1175.5	24.7	1070	1	AGLU_CANTS
19	978	20.5	923	1	AGLU_TETPY
20	956.5	20.1	787	1	AGLU_BACTQ
21	787	16.5	954	1	YB79_YEAST
22	676	14.2	693	1	AGLU_SULSO
23	519.5	10.9	731	1	XYLS_SULSO
24	458	9.6	772	1	YICI_ECOLI
25	301.5	6.3	678	1	YIHQ_ECOLI
26	231.5	4.9	529	1	YCR2_ERWHE
27	158	3.3	696	1	SP15_TORCA
28	129.5	2.7	737	1	SKN1_CANAL
29	127	2.7	2477	1	FINC_RAT
30	124.5	2.6	2386	1	FINC_HUMAN
31	124.5	2.6	3354	1	CADN_MOUSE
32	122	2.6	3124	1	CA1C_CHICK
33	120.5	2.5	1284	1	NRCA_CHICK

34	120	2.5	3354	1	CADN_HUMAN
35	119.5	2.5	3317	1	CADN_RAT
36	116	2.4	719	1	AGA2_PEDPE
37	116	2.4	733	1	AGAI_PEDPE
38	114	2.4	1967	1	CD87_DROME
39	113	2.4	982	1	POL_HTLV2
40	111.5	2.3	518	1	VL2_HPV36
41	110.5	2.3	467	1	SMA9_HUMAN
42	110	2.3	468	1	PGD_BUCBP
43	109.5	2.3	557	1	PGMU_EMENI
44	109.5	2.3	1170	1	CIJB_BACTU
45	109.5	2.3	3063	1	CA1C_HUMAN

Q9h251 homo sapien  
P58365 rattus norv  
P43469 pedicoccus  
Q9v995 drosophila  
P03363 human t-cel  
P50827 human papil  
O15198 homo sapien  
Q89ax5 buchnera ap  
Q9p931 emericeila  
Q45716 bacillus th  
Q99715 homo sapien

## ALIGNMENTS

### RESULT 1

AGLU\_SPTOL  
ID AGLU\_SPTOL STANDARD; PRT; 903 AA.  
AC O04893;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase).  
OS Spinacia oleracea (Spinach).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.  
OX NCBI\_TaxID=3562;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=cv, Dash;  
RX MEDLINE=97238484; PubMed=9132069;  
RA Sugimoto M., Furui S., Suzuki Y.;  
RT "Molecular cloning and characterization of a cDNA encoding alpha-glucosidase from spinach."  
RJ Plant Mol. Biol. 33:765-768(1997).  
CC -!- FUNCTION: ALPHA-GLUCOSIDASE I AND II HAVE HIGH ACTIVITY TOWARDS MALTO-OLIGOSACCHARIDES AND STARCH, WHILE FORM III AND IV HAVE HIGH ACTIVITY TOWARDS MALTO-OLIGOSACCHARIDES BUT LOW ACTIVITY TOWARD STARCH.  
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-linked D-glucose residues with release of D-glucose.  
CC -!- PTM: FOUR DIFFERENT FORMS (I-IV) MAY BE PRODUCED BY POST-TRANSLATIONAL MODIFICATION.  
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL: D86624; BAA19924.1; --  
CC PIR: T09143; T09143.  
CC InterPro: IPR000322; Glyco\_hydro\_31.  
CC Pfam: PF01055; Glyco\_hydro\_31; 1.  
CC PROSITE: PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
CC PROSITE: PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.  
CC Hydrolase; Glycosidase; Glycoprotein; Signal.  
CC SIGNAL 1 24 POTENTIAL.  
CC CHAIN 25 903 ALPHA-GLUCOSIDASE.  
CC ACT SITE 465 465 BY SIMILARITY.  
CC CARBOHYD 64 64 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC CARBOHYD 219 219 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC CARBOHYD 400 400 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC CARBOHYD 491 491 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC CARBOHYD 600 600 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC SEQUENCE 903 AA; 100880 MW; 58054E27C20EC33A CRC64;

Query Match		100.0%	Score 4768;	DB 1;	Length 903;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 903;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MKKKIFSLALGILLVFLQYLVAGISTSENDEGVGYGVKVKVDSTGTRSLTA	PQ 60		
Db	1	MKKKIFSLALGILLVFLQYLVAGISTSENDEGVGYGVKVKVDSTGTRSLTA	PQ 60		
Qy	61	LVQNSVYGPDIQLLSITASLENDRLVRITDAKHRRWEIPDNILHRHQPPPPHSL	120		
Db	61	LVQNSVYGPDIQLLSITASLENDRLVRITDAKHRRWEIPDNILHRHQPPPPHSL	120		
Qy	121	SLVRTLSSPTTNRKILLSHPNSDLTFSINTTPEFTISRKSTHDVLFDPDTPN	180		
Db	121	SLVRTLSSPTTNRKILLSHPNSDLTFSINTTPEFTISRKSTHDVLFDPDTPN	180		
Qy	181	TLFIDQYLHLTSSLPGRTRAHYIHLGHSKPTFQLAHNQTLTMRAADIPSSNPOV	240		
Db	181	TLFIDQYLHLTSSLPGRTRAHYIHLGHSKPTFQLAHNQTLTMRAADIPSSNPOV	240		
Qy	241	SHPFYMDVRSSPVAGSTHGVLLNSNGMDVEYTCGNRITYKVGIIIDLYFFAGP	300		
Db	241	SHPFYMDVRSSPVAGSTHGVLLNSNGMDVEYTCGNRITYKVGIIIDLYFFAGP	300		
Qy	301	EQFTRVIGRPAFQCRQYGVHDVYELQSVVAGYAKAKIPLVWMTDIDYMDAY	360		
Db	301	EQFTRVIGRPAFQCRQYGVHDVYELQSVVAGYAKAKIPLVWMTDIDYMDAY	360		
Qy	361	KDFTLDVNPFLDKMKFVNHLHNGQKYYVILDPGISTNKTYETIRGMKHDFV	420		
Db	361	KDFTLDVNPFLDKMKFVNHLHNGQKYYVILDPGISTNKTYETIRGMKHDFV	420		
Qy	421	KPYLGSVWPCPVFPDPLKPSALTMTDEIKRFLNLLPVDGLMIDNNEISFI	480		
Db	421	KPYLGSVWPCPVFPDPLKPSALTMTDEIKRFLNLLPVDGLMIDNNEISFI	480		
Qy	481	STLDNPPYKINNSGVMLPIINKTIPPTAMHYGDIPEYVNHVLFGLYEA	540		
Db	481	STLDNPPYKINNSGVMLPIINKTIPPTAMHYGDIPEYVNHVLFGLYEA	540		
Qy	541	KRPFLVSRSTFGSGKYTAHTWGDNAATWMDLVYSIPSLDFGLFGIPMV	600		
Db	541	KRPFLVSRSTFGSGKYTAHTWGDNAATWMDLVYSIPSLDFGLFGIPMV	600		
Qy	601	TTTELCRMWQLGAFYFSDHSLGTYOELRWESVAASARKVILGLRVTLL	660		
Db	601	TTTELCRMWQLGAFYFSDHSLGTYOELRWESVAASARKVILGLRVTLL	660		
Qy	661	YEAQLNGIPIARPLFFSPDDIKTYGSSQFLGKGMVSPVLKPGVSVTAY	720		
Db	661	YEAQLNGIPIARPLFFSPDDIKTYGSSQFLGKGMVSPVLKPGVSVTAY	720		
Qy	721	DLFDYTSVNTASTGRYVTLSPADPHINVHIOENILAMQKANTOAR	780		
Db	721	DLFDYTSVNTASTGRYVTLSPADPHINVHIOENILAMQKANTOAR	780		
Qy	781	DCGASFGLEFLDDGVETMGVGRKMTFVKAFIAASAKOTCIITSDVVS	840		
Db	781	DCGASFGLEFLDDGVETMGVGRKMTFVKAFIAASAKOTCIITSDVVS	840		
Qy	841	KVTILGRKTKINGYVTRTGAVTRKGDKSLKSTPDRKGEFIVAEISGL	900		
Db	841	KVTILGRKTKINGYVTRTGAVTRKGDKSLKSTPDRKGEFIVAEISGL	900		
Qy	901	VLH 903			
Db	901	VLH 903			

RESULT 2

AGLU\_BETVU

ID \_AGLU\_BETVU STANDARD; PRT; 913 AA.

AC	004931;				
DT	15-JUL-1998 (Rel. 36, Created)				
DT	15-JUL-1998 (Rel. 36, Last sequence update)				
DT	30-MAY-2000 (Rel. 39, Last annotation update)				
DE	Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase).				
OS	Beta vulgaris (Sugar beet).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;				
OC	Caryophyllales; Caryophyllaceae; Chenopodiaceae; Beta.				
NCBI	taxid=161934;				
QY	(1)				
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				
RC	STRAIN=CV, NK-152;				
RX	MEDLINE=97321863; PubMed=9178565;				
RA	Matsui H., Iwanami S., Ito H., Mori H., Honma M., Chiba S.;				
RT	"Cloning and sequencing of a cDNA encoding alpha-glucosidase from				
RT	sugar beet.";				
RL	Biochem. Biotechnol. Biochem. 61:875-880(1997).				
RP	(2)				
RP	ACTIVE SITE, AND SEQUENCE OF 464-472.				
RX	MEDLINE=9525292; PubMed=7766184;				
RA	Iwanami S., Matsui H., Kimura A., Ito H., Mori H., Honma M.,				
RA	Chiba S.;				
RT	"Chemical modification and amino acid sequence of active site in				
RT	sugar beet alpha-glucosidase.";				
RL	Biochem. Biotechnol. Biochem. 59:459-463(1995).				
CC	!- FUNCTION: HIGH ACTIVITY FOR ALPHA-GLUCAN.				
CC	!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-				
CC	linked D-glucose residues with release of D-glucose.				
CC	!- PTM: THE N-TERMINUS IS BLOCKED.				
CC	!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL; D89615; BAA20343.1; ..				
DR	PIR; JC5463; JC5463.				
DR	InterPro; IPR000322; Glyco hydro_31.				
DR	Pfam; PF01055; Glyco hydro_31; 1.				
DR	PROSITE; PS00129; GLYCOSYL HYDROL F31_1; 1.				
DR	PROSITE; PS00707; GLYCOSYL HYDROL F31_2; 1.				
KW	Hydrolase; Glycosidase; Glycoprotein; Signal.				
FT	SIGNAL 1 28 POTENTIAL.				
FT	CHAIN 29 913 ALPHA-GLUCOSIDASE.				
FT	ACT SITE 469 469				
FT	CARBOHYD 54 54 N-LINKED (GLCNAC. .) (POTENTIAL).				
FT	CARBOHYD 404 404 N-LINKED (GLCNAC. .) (POTENTIAL).				
FT	CARBOHYD 495 495 N-LINKED (GLCNAC. .) (POTENTIAL).				
FT	CARBOHYD 517 517 N-LINKED (GLCNAC. .) (POTENTIAL).				
FT	CARBOHYD 728 728 N-LINKED (GLCNAC. .) (POTENTIAL).				
FT	CARBOHYD 823 823 N-LINKED (GLCNAC. .) (POTENTIAL).				
QY	SEQUENCE 913 AA; 102117 MW; 02AC4F0B505369CC CRC64;				
Query Match					
Best Local Similarity		68.0%;	Score 3243.5;	DB 1;	Length 913;
Matches 610;		Conservative 122;	Mismatches 166;	Indels 9;	Gaps 5;
Qy	2	KKKIPSL---ALGILLVFLQYLVAGISTSENDEGVGYGVKVKVDSTGTRSLTA	57		
Db	3	RSKLPRYICPTLAVVLPVLVLCMVVEGATTCKNDNGEAGVGYGVKVDSTGKSLTA	62		
Qy	58	LPOLVQNSVYGPDIQLLSITASLENDRLVRITDAKHRRWEIPDNILHRHQPPPPH	117		
Db	63	LLQLIRNSPVYGPDIHFLSEFASFEEDDTLRIRFDANRRRWEIPDNVLP-PPPPSP	121		
Qy	118	SLSSILYRTLSSPTTNRKILLSHPNSDLTFSINTTPEFTISRKSTHDVLFDPDTP	177		
Db	122	PLSSQLHLPKPIPONQPTTTVLSPHSDLAFTLPHFTTIFRSTHDLVDFDTP	181		

```

QY 178 NPNTLFIIDYVLTSSLPGRTRAHYGLGHSKPTFQLAHQNTLTWRAADIPSSNDPVN 237
DB 182 NPTTLLIKDQVQLSSSLPAQAAHLYGLGHTKPTFQLAHQNTLTWNAADIASFNEDLN 241
QY 238 LYGSHFFMDVRSSPVAGSTHGVLLNSGMDVETGNRIYKVIGGIIDLYFFAGSPG 297
DB 242 LYGSHFFMDVRSSPMVSGTHGVFLNSGMDVETGNRIYKVIGGIIDLYIFAGRTPE 301
QY 298 QVVEQFTVIGRPAPEYMAFGVQOCRYGHDVYELQSVVAGYAKAKIPLVEMMTDIDYM 357
DB 302 WYLDQYTKLIGRAPMPYAFGHOCRYGHDVYELQSVVAGYAKAKIPLVEMMTDIDYM 361
QY 358 DAYKDFTLDPVNFPLDKMKKFNVLHKNQKQYVWILDFGISTNKTYYIIRGMKHDVFLK 417
DB 362 DAFKDFTLDPVNFPLDKMQQVFTKLHNGQRYVPILDPGINTKNSYGTFFIRGMQSNVFIK 421
QY 418 RKGKYLGSVWPGVYFDFLKPALSALTWTWDEIKRFLNLLPVDGLWIDMNEISFISPP 477
DB 422 RGNPNYLGVSWPGVYFDFLKPALSALTWTWDEIKRFLNLLPVDGLWIDMNEISFISAP 481
QY 478 IPGSTLDNPPYKINNSGVMLPIINTIPTTAMHYGDIPEYVNVHNLFGYLEARVTRAALIK 537
DB 482 TFGSTLDNPPYKINNSGGRVPINSKTIIPATAMHYGNVTEYNAHNLXGLESQATREALVR 541
QY 538 LTKPKPFLVSRSTFGSGKGYTAHTWTDGNAATWDLVYISIPMLDFGLFGIPMGADICGF 597
DB 542 PATRGPFLLSRSTFAGSGKGYTAHTWTDGNAATWDLVYISIPMLDFGLFGIPMGADICGF 601
QY 598 LONTTEELCRRIOLGAFYFSDRSHSLGTTVQELRYHESVAASARKVLGRYTLIPYFY 657
DB 602 AESTTEELCRRIOLGAFYFSDRSHSLGTTVQELRYHESVAASARKVLGRYTLIPYFY 661
QY 658 TLMEYEAQNGIPIARPLFFSPDDIKTYGSSQFLLGKGMVSPVLKPGVWVSATYFPRG 717
DB 662 TLXYDANLGSPIARPLSFTFDDVATYGISQFJIGRGIWSPVLQPGSSIVNAYSFRG 721
QY 718 NWFDLFDYTRVSTAGRYVTLUSAPDHNHVIQEGNITAMOGKAMTTOAARKTPFHLLV 777
DB 722 NWVLSNVTSSVSAGTYVLSLAPPDHNHVIHEGNTVAVQCEAMTTOAARSTPFHLLV 781
QY 778 VMSDCGASGELFDLGDGVMTGVNVRGKWTFTKPTAASAKOTCIITSDVSGEFAVSQKW 837
DB 782 VMSDHVASTGELFDLGDGVMTGVNVRGKWTFTKPTAASAKOTCIITSDVSGEFAVSQKW 841
QY 838 VIDKVTILGLRGTGKINGYTVR - TGAVTRKGDKSCLKASTPRKGEFTVABISGLKLLG 895
DB 842 VMDKITILGLRGTGKINGYTVR - TGAVTRKGDKSCLKASTPRKGEFTVABISGLKLLG 895
QY 896 REFKLVL 902
DB 900 CAFKLEL 906

```

RESULT 3

AGLU\_HORVU STANDARD; PRT; 877 AA.

AC Q43763;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase).

OS Hordeum vulgare (Barley).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;

OC Triciceae; Hordeum.

OX NCBI\_TaxID=4513;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Morex; TISSUE=Aleucone;

RX MEDLINE=9617863; PubMed=8616248;

RA Tibbot B.K., Skadsen R.W.

RT "Molecular cloning and characterization of a gibberellin-inducible,

```

RT putative alpha-glucosidase gene from barley.";


RL Plant Mol. Biol. 30:229-241(1996).



CC -i- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-



CC linked D-glucose residues with release of D-glucose. AND SCUTELLUM



CC -i- TISSUE SPECIFICITY: HIGH LEVELS SEEN IN THE ALEURONE AND SCUTELLUM



CC AFTER GERMINATION, WHILE LOW LEVELS ARE FOUND IN DEVELOPING SEEDS.



CC -i- DEVELOPMENTAL STAGE: LEVELS INCREASE STEADILY THROUGHOUT



CC IMBIBITION REACHING MAXIMUM LEVELS AT DAY 7. DURING GERMINATION,



CC LEVELS INCREASE FROM DAY 2, REACH MAXIMUM LEVELS AT DAY 3 AND



CC DECLINE AFTER DAY 5.



CC -i- INDUCTION: BY GIBBERELLIN A3 (GA).



CC -i- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.



CC This SWISS-PROT entry is copyright. It is produced through a collaboration



CC between the Swiss Institute of Bioinformatics and the EMBL Outstation



CC the European Bioinformatics Institute. There are no restrictions on its



CC use by non-profit institutions as long as its content is in no way



CC modified and this statement is not removed. Usage by and for commercial



CC entities requires a license agreement (See http://www.isb-sib.ch/announce/



CC or send an email to license@isb-sib.ch).



CC -----



DR EMBL: U22450; AAC02985.1; -



DR PIR: S65057; S65057



DR InterPro: IPR000322; Glyco_hydro_31.



DR Pfam: PF01055; Glyco_hydro_31; 1.



DR PROSITE: PS00129; GLYCOSYL_HYDROL_F31_1; 1.



DR PROSITE: PS00707; GLYCOSYL_HYDROL_F31_2; 1.



KW Hydrolase; Glycosidase; Glycoprotein; Signal.



KW SIGNAL 1 23 POTENTIAL.



FT CHAIN 24 877 ALPHA-GLUCOSIDASE.



FT ACT_SITE 437 437 BY SIMILARITY.



FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).



FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (POTENTIAL).



FT CARBOHYD 338 338 N-LINKED (GLCNAC. . .) (POTENTIAL).



FT CARBOHYD 391 391 N-LINKED (GLCNAC. . .) (POTENTIAL).



FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).



FT CARBOHYD 570 570 N-LINKED (GLCNAC. . .) (POTENTIAL).



SQ SSEQUENCE 877 AA; 96933 MW; E3B5C16A598C492 CRC64;



Query Match 50.8%; Score 2422.5; DB 1; Length 877;



Best Local Similarity 53.6%; Pred. No. 5.4e-162;



Matches 486; Conservative 132; Mismatches 234; Indels 55; Gaps 15;



Qy 10 LGILLVFLVQLVAG--ISTNDP---EGVIGYGVYK-SVKVDSGT-RRSLTALPQLV 62



Db 4 VGVLLCLCLCLFAPRLCSSKEGPLAARTVLAVALMEGALRAEAATGCRSGT----- 58



Qy 63 KNSVVGPIQLLSITASLESNDRLRVRTDAKRRWEIPDNLHRHQ-----PPPPPHS 118



Db 59 -----DVQRLAVYASLETSLRVRTDAKRRWEIPDNLHRHQ-----PPPPPHS 110



Qy 119 LSSLVYRTLLSSPTNRRKILLSHPNSDLTFSINTTTPFGFTISRSKSTHVDLFDATPDPTN 178



Db 111 -----SAPLQGR---VLSPGSDLVLT-VHASPFTVSRSTGTDLFDATPG--- 154



Qy 179 PNTFLIFDIQYHLTSSLPGRTRAHYVGLGHSKPTFQLAHQNTLTWRAADIPSSNDPVN 238



Db 155 -----LVFRDKYLEVTSALPAGRASLYGLGHTKSSFLRHNSFTLWNADIGASYVDNL 210



Qy 239 YGSHPPFMDVRSSPVAGSTHGVLLNSGMDVETGNRIYKVIGGIIDLYFFAGSPG 298



Db 211 YGSHPPFMDVRA---PGTAHGVLLNSGMDVLYGGSYVTVYKVIGVLDLYFFAGNP 267



Qy 299 VVQFTRVIGRPAPEYMAFGVQOCRYGHDVYELQSVVAGYAKAKIPLVEMMTDIDYM 358



Db 268 VVDQYQLARPAPEYMAFGVQOCRYGHDVYELQSVVAGYAKAKIPLVEMMTDIDYM 327



Qy 359 AYKDFTLDPVNFPLDKMKKFNVLHKNQKQYVWILDFGISTNKTYYIIRGMKHDV 415



Db 328 GFKDFTLDRVNFPAELRPFDLKPALSALTWTWDEIKRFLNLLPVDGLWIDMNEIS 387



Qy 416 LKGNKPYLGVSWPGVYFDFLKPALSALTWTWDEIKRFLNLLPVDGLWIDMNEIS 475


```

```

DR 388 LKRNFTNFVGNWPGDVTFDFMFHAAEFNAREISLFRRTIPVDGLWDMNEISFN- 446
QY 476 PPIPOSTLNDPPYKINNNGWMLPIINKTIPPTAMHYGDIPEYNVHNLPGLYLEAVTRAAAL 535
DB 447 -PEPNALDDPPYRINNDDTGRPINNKTRPLAVHYGGVTEYEENLFGLEAARATGRV 505
QY 536 IKLTEKRPFLVLSRSTFSGSGKYTAHWTGDNATWDLVSPSMLDFGLFGIPMVGDADIC 595
DB 506 LRDTRRRPFLVLSRSTFVSGRYTATWGNATWGLDLYSNTWLSFGFLGPNMIGADIC 565
QY 596 GFLGNTTELCRRWITQLGAFYFSDHSSLGTTYOELYRWESVAASARKVLGLRYTLPLY 655
DB 566 GPNGNTTELCGRWITQLGAFYFSDHSSAIFTVRRELYLMPVSAASGRKALGLAYQLLPY 625
QY 656 FYTLMEYEAQNCIPARLFFSPFDDIITYGISSQFLLGKGMVSPVLKPGVSVTAYFP 715
DB 626 FYTLMEYEAHMTGAPARLFFSYPHDVATYGVDRQFLGRLGVLPVLEPGFTVDAYFP 685
QY 716 RGNWFLDFYTRSVTASTGRVYVTLSPDPHINVH:QEGNILAMQGMKMTTCAARKTFFHL 775
DB 686 AGRWRRLDYSLAVATRGKVRLPAPADTVNVHLTGGTILPLOSALTTSRARTAFHL 745
QY 776 LVVMSDCGASGFEFLDDGVEVTGMVGRKMTFVKF--IAASAKQTCITSDVSGGEFV 833
DB 746 LVALLAEDGTASGYFLDDGDSPEYG-RRSDMSVRFNYKIPNNKGAIKVKSEVHNSYAO 804
QY 834 SOKWIDKVTILGRKGTINKINGTYVTRGAVTRKGDKSLKSTPPDKGFEFVAEISGLNLL 893
DB 805 SRTLVISKVLVLMGHRSPAPKKLTVHNSAEVEASSAGTRYQYAGGLGGVVAHIGGSLV 864
QY 894 LGREFKL 900
DB 865 VGEFEEL 871

RESULT 4
AGLU MUCJA
ID AGLU MUCJA STANDARD; PRT; 864 AA.
AC Q92442;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase).
OS Mucor javanicus.
OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
OC Mucor.
OC NCBI_TaxID=51122;
RN (1)
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=IFO 4570;
RX MEDLINE=96271012; PubMed=8830045;
RA Sugimoto M., Suzuki Y.;
RT "Molecular cloning, sequencing, and expression of a cDNA encoding
alpha-glucosidase from Mucor javanicus.";
RL J. Biochem. 119:500-505(1996).
CC !- FUNCTION: Hydrolyzes not only malto-oligosaccharides but also
soluble starch.
CC !- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-
linked D-glucose residues with release of D-glucose.
CC !- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: D67034; BAA11053.1;
CC InterPro: IPR000322; Glyco_hydro_31.
CC Pfam: PF01055; Glyco_hydro_31.1.
CC PROSITE: PS00129; GLYCOSYL_HYDROL_F31_1; 1.

```

```

DR PROSITE: PS00707; GLYCOSYL_HYDROL_F31_2; FALSE_NEG.
KW Hydrolase; Glycosidase; Glycoprotein; Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 864 ALPHA-GLUCOSIDASE.
FT ACT_SITE 430 430 BY SIMILARITY.
FT CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 364 364 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 406 406 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 466 466 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 500 500 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 734 734 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 864 AA; 98761 MW; CFAB4759DC431403 CRC64;

Query Match 31.4%; Score 1499; DB 1; Length 864;
Best Local Similarity 38.2%; Pred. No. 3.2e-97;
Matches 324; Conservative 147; Mismatches 280; Indels 98; Gaps 18;

QY 12 ILLVFLQLVLVAGISTSENDEPGVIGY--GKVKV--VKVDSGTRRSRLTALPQLVKNSS 66
DB 4 VSFIFAIALITGNVLCQTDATYAVSSAPGYKIDGHVRKTEAGLHPLTLNSRGNKKTG 63
QY 67 V--YGPDIQLLSITASLENDRLRVRITDAKHRRWEIPDNILHRHQPPPPPPHSLSLYR 124
DB 64 IDTFGKTIDITVDVEYETEERLHVKISKAKKQYLPVDSPLGFERP----- 110
QY 125 TLSSPTTNRKILLGHPNSDLTFLSINTTFFGTTISRKSTHDVLFDATPDPTNPNTFLI 184
DB 111 -----QIKHYVSPKHSNLDFO-QYAKPFSGKVRKDDKTTIIFDTNMP-----LV 154
QY 185 FIDQYLHLTSSLPSTRAHYGLGEHSKPTFQLAHNOTLTMRADIFSSNPD---VNLYGS 241
DB 155 FEQYLELSTKVP-EDANIYIGEVETAP-FRRTHNVT-TLWARD---NPDDFYRNIYA 207
QY 242 HPFMDVRSSPVAGSTHGULLNSGMDVEYTGRIYKYVIGIILYFFAGS--PGQV 299
DB 208 HPYYOEVRD---GKAHGALLANAHGMDVITTEGRITKYVIGGILDFYFAPKSGKPNDL 263
QY 300 VEOFTRVIGRPAMPYMAFGFQCRYGHDVYELQSVAGYAKAKIPLVEMWTDIDYMDA 359
DB 264 SIATYDLIGKPMPSHMLGMWHCRIGYPNIDKVEYKVKYKEANIPLOTWVDIDYME 323
QY 360 YKFTLPDPVNFPLDKMKKFFVNNLHKQKQYVILDFGISTNKTETTYIRGMKHDVFLKR- 418
DB 324 TKDFTFKVNFPODRMIGLGEQLHKDQYVVMVDPASISANTTYEPVVRGTENDVMIKNA 383
QY 419 NGKPYLGSVWPGVPYVDFLKPALSALTFTDEIRKFNLLPVDGLWDMNEISNF----- 472
DB 384 DGSDFIGSVWPGFTTFPDMWHPNATKYMNKEIIDFVDMLGVDGLWDMNEFASFCILGSCG 443
QY 473 -----ISSPPIFGSLDNPYPKINNSGVMLPI 499
DB 444 SGKVDAGNQPYRWYTYEEEOAANHTRWEKELKAMGNPPGGEERNLLYPKYAINNAGNLS- 502
QY 500 INKTIPPTAMHYGDIPEYNVHNLPGLYLEAVTRAAIKLITEK-RPFVLSRSTSGSKYT 558
DB 503 -EFTVATTALHYGNIPHYDHNLYGHAESHITROALIKHKNKIRPFVLTSSPGSGKSV 561
QY 559 AHWTDGNAATWDLVYSIPSMDFGLFGIPMVGDADICGFLGNTTELCRRWITQLGAFYFP 618
DB 562 GHWTGDNHSPWYLPKNSIANILNFMQFVSGVSGADVCGFNSDTTELCRTWMEIGAFYFP 621
QY 619 SRDSSSLGTTYOELYRWESVAASARKVLGLRYTLFPYVYTLMEYEAQNGIPIARPLFFSF 678
DB 622 ARNNNNAADQEPYLNWSTAEASRIANTRYEMLPYFYLFEESNRLGLGWRPLIFEY 681
QY 679 PDDIKTYGISQSQFLGKGMVSPVLKPGVSVTAYFPRGNWFDLFDYTRSVTASTG--R 735
DB 682 PAYEELVNDVQTLVSGDILLSPVLDSGKTSVKQAQFPGQWYDWTYHETLVTDNKNKKVK 741
QY 736 YVTLSAPDPHINVHIQEGNILAMQGMKMTTQAAKRTPFHLLVMSDCGASGFEFLDDGV 795
DB 742 TVTLDAPLTHIPIHIGGAIIPTKPKYTVGETFATPYNLVIALDKKQGOASGRLYIDDGE 801

```

QY 796 EVTMGVNRG 804  
Db 802 SLEVKSSG 810

## RESULT 5

LYAG MOUSE STANDARD; PRT; 953 AA.  
AC P70699;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Lysosomal alpha-glucosidase precursor (EC 3.2.1.20) (Acid maltase).  
GN GAA.  
OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=BA1B/c; TISSUE=Liver;

RA Ding J.H., Yang B.Z., Reuser A.J.J., Roe C.R.;  
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.

CC FUNCTION: ESSENTIAL FOR THE DEGRADATION OF GLYCOGEN TO GLUCOSE IN

CC LYSOSOMES.

CC CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-

CC linked D-glucose residues with release of D-glucose.

CC SUBCELLULAR LOCATION: Lysosomal.

CC SIMILARITY: Contains 1 P-type (trefoil) domain.

CC SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U49351; AB06943.1; ..

DR HSP; P01359; 2PSP.

DR MGD; MGI:95609; Gaa.

DR InterPro; IPR000322; Glyco\_hydro\_31.

DR InterPro; IPR000519; P\_trefoil.

DR Pfam; PF01055; Glyco\_hydro\_31; 1.

DR Pfam; PF00088; trefoil; 1.

DR SMART; SM00018; PD; 1.

DR PROSITE; PS00025; P\_TREFOIL; 1.

DR PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.

DR PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.

KW Hydrolase; Glycosidase; Glycoprotein; Lysosome; Signal.

FT SIGNAL 1 27 POTENTIAL.

FT PROPEP 28 69 BY SIMILARITY.

FT CHAIN 70 953 LYSOSOMAL ALPHA-GLUCOSIDASE.

FT DOMAIN 81 130 P-TYPE.

FT ACT\_SITE 518 518 BY SIMILARITY.

FT DISULFID 82 109 BY SIMILARITY.

FT DISULFID 92 108 BY SIMILARITY.

FT DISULFID 103 127 BY SIMILARITY.

FT CARBOHYD 140 140 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 390 390 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 883 883 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 926 926 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 933 933 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SEQUENCE 953 AA; 106168 MW; 927EE0C7B1D59571 CRC64;

Query Match

Best Local Similarity 31.4%; Score 1498; DB 1; Length 953;

Matches 336; Conservative 136; Mismatches 317; Indels 98; Gaps 23;

QY 40 YKVSXKVDSTGRSLTALPOLVKNSVVGPR-DIOLLSITASLESNDRLRVIITDAKRR 98  
Db 136 YRLNDL---SSTESGYTA--TLTRISPTFFKDVLTLOLEVMETDSRHKFKIKDPASKR 190  
QY 99 WEIPDNILHRHQPPPPPHSLSSLYRTLSSPTTNRRLKILLSHPNSDLTFLSINTPTPGF 158  
Db 191 YEVP-----LETPR-----VLSQAFSPLYSVFEFSEPGV 220  
QY 159 TISRKSTHVDLPDPTDPTNPTPLIFIDQYHLTSSLPQTRAHYIYGLGEHSKPTFQLAH 218  
Db 221 IYRRKLGGRVLLNTTVP-----LFFADQFLQLSTLPA--QHTISLGLHSLPLMLSTD 272  
QY 219 NQTLTMRADIPSSNPDPVNLGSHHPYMDVRSSPVAGSTHGVLLNSGMDVETGN-RI 277  
Db 273 WARIILNRDTPSQ-GTNLGSHPYLALED---GGLAHGVFLNSAMVDVILQPSAL 328  
QY 278 TYKVGIIIDLYFRAGSPGQVQVEQTRVIGRPAPMPYAFGFOOCRYGYHDVVEQSVW 337  
Db 329 TWRSTGGILDVTVFLGPEPKSVVQOYLDVVGVPFPPYWGVLGFLCRWGSYSTAIVQV 388  
QY 338 AGYAKAKIPEVMWTDIDYMDAYKDTLDVNFPLDKMKKFNHLHKNQKVVVILDPGI 397  
Db 389 ENWTRTHFFLDVQVNDLDYMDARRDFTFNQDSFA--DFPDVVRDVHGGRRYMMIVDAI 446  
QY 398 ST---NKTYETIYINGMCHDVFL-KENGKPYLGSMVPGVYFPDFLKPASLTFTWDEIKRF 453  
Db 447 SSAGPAGSVRPYDEGLRGVFTNETGQPLIGKVCPTTAPDPTNPETLDWQDMVSEF 506  
QY 454 LMLLPVDGLWIDMNEISNFI---SSPPIPGSTLDNPPYKINNSGVMLPIINKTIPPTAMHY 511  
Db 507 HXQVPPDGMMDMNEPSNFVRGSGQCCPNNELENPPYV---PGVVGGLQATICASSHQ 563  
QY 512 GDPIPEYVNLHGLYLEARVTRAALIKLTKRPFVLSRSTFGSGKYTAHTWCDNAATWND 571  
Db 564 FLSTHNLHNLGLTEATASSALVKTRCTRPVLSRSTFGSGHYAGHWEDVRTSWEH 623  
QY 572 LVYSTPSMLDGLGIPVWGADICGFLGNTTTELCELRMIQLGAFYPPFRDSSSLGTTVOE 631  
Db 624 LAYSVDIILQFNLLGVPLVGADICGFIGDTSBELCVRTQLGAFYPPFRMNRHNDLNSVPQE 683  
QY 632 LYRWSVAASA-RKVLGLRYTLTPYFTLMEYLAQLGPIARPLFFSPDDIKTVGISQ 690  
Db 684 PYRSETAQANMRKAPALRYALLPYTLPHRAVRGDTVARPLFLEFREDPSTMSVDRQ 743  
QY 691 FLGLGGMVSPVLKPGVSWSTAYFPFRGNWFDL-----FDYTRSVTASTG 734  
Db 744 LLWGPALLITPVLGKTEVTGTFPKGTWYNNQVVSVDLGLTLPSPSSASSFRSAVQSKG 803  
QY 735 RYVTLGAPPDHINHIQEGNIIAMOGKAMTQAARKTPFHLVVMDCGASFGELFLDDG 794  
Db 804 QMLTLEAPLDTINHLREGYIIPQGPSLTTTTSRKQPMALAVALTASGEADGELFWDG 863  
QY 795 VEVTVGNRGKWTFFVKAASAKQTCIITSV-VSGEFAVSQKVVIDKVTILGLRKGTKI 853  
Db 864 ESLAV-LEHANTLVTF---SANNNTIVKLVVRVTKEGAELO---LKEVTVLGVA----- 911  
QY 854 NGYVTRTGAVTRKDGKSLKSTPDRKGEFIVAEISGLNLLGLREFKL 900  
Db 912 ---TAPTQVLSNGIPVSNFTYSPDNKSLAI-----PVSLLMGELFQI 950

## RESULT 6

LYAG HUMAN

ID LYAG HUMAN STANDARD; PRT; 952 AA.

AC P10253; Q14351; Q16302;

DT 01-MAR-1989 (Rel. 10, Created)

DT 01-FEB-1991 (Rel. 17, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Lysosomal alpha-glucosidase precursor (EC 3.2.1.20) (Acid maltase).

GN GAA.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.



RX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE OF 70-89; 123-145; 204-215;  
 RA 230-249; 332-345; 349-370; 394-409; 480-513; 520-545; 703-719; 726-731  
 RP AND 795-803.  
 RC TISSUE=Placenta, Testis, and Urine;  
 RX MEDLINE=89005058; PubMed=3049072;  
 RA Hofstloot L.H., Hooijveen-Westerveld M., Kroos M.A., van Beunen J.,  
 RA Reuser A.J.J., Oostra B.A.;  
 RT "Primary structure and processing of lysosomal alpha-glucosidase;  
 RT homology with the intestinal sucrase-isomaltase complex.";  
 RL EMBO J. 7:1697-1704(1988).  
 RN [2]  
 RP REVISIONS.  
 RA Reuser A.J.J.;  
 RL Submitted (JUN-1990) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90262651; PubMed=2111708;  
 RA Martinik F., Mehler M., Tzall S., Meredith G., Hirschhorn R.;  
 RT "Sequence of the cDNA and 5'-flanking region for human acid alpha-  
 RT glucosidase, detection of an intron in the 5' untranslated leader  
 RT sequence, definition of 18-bp polymorphisms, and differences with  
 RT previous cDNA and amino acid sequences.";  
 RL DNA Cell Biol. 9:85-94(1990).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91097465; PubMed=2268276;  
 RA Hofstloot L.H., Hooijveen-Westerveld M., Reuser A.J.J., Oostra B.A.;  
 RT "Characterization of the human lysosomal alpha-glucosidase gene.";  
 RL Biochem. J. 272:493-497(1990).  
 RN [5]  
 RP ACTIVE SITE.  
 RX MEDLINE=91310614; PubMed=1856189;  
 RA Hermans M.M.P., Kroos M.A., van Beunen J., Oostra B.A.,  
 RA Reuser A.J.J.;  
 RT "Human lysosomal alpha-glucosidase. Characterization of the catalytic  
 RT site.";  
 RL J. Biol. Chem. 266:13507-13512(1991).  
 RN [6]  
 RP CARBOHYDRATE-LINKAGE SITES.  
 RX MEDLINE=93168114; PubMed=8435067;  
 RA Hermans M.M.P., Wisselaar H.A., Kroos M.A., Oostra B.A.,  
 RA Reuser A.J.J.;  
 RT "Human lysosomal alpha-glucosidase: functional characterization of  
 RT the glycosylation sites.";  
 RL Biochem. J. 289:681-686(1993).  
 RN [7]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=95327152; PubMed=7603530;  
 RA Reuser A.J.J., Kroos M.A., Hermans M.M.P., Bijvoet A.G.A.,  
 RA Verbeet M.P., van Diggelen O.P., Kleijer W.J., van der Ploeg A.T.;  
 RT "Glycosenosis type II (acid maltase deficiency).";  
 RL Muscle Nerve 13:S61-S69(1995).  
 RN [8]  
 RP VARIANT ASN-91.  
 RX MEDLINE=90365036; PubMed=2203258;  
 RA Martinik F., Bodkin M., Tzall S., Hirschhorn R.;  
 RT "Identification of the base-pair substitution responsible for a human  
 RT acid alpha glucosidase allele with lower 'affinity' for glycogen (GAA  
 RT 2) and transient gene expression in deficient cells.";  
 RL Am. J. Hum. Genet. 47:440-445(1990).  
 RN [9]  
 RP VARIANT GSD-II THR-318.  
 RX MEDLINE=91353580; PubMed=1652892;  
 RA Zhong N., Martinik F., Tzall S., Hirschhorn R.;  
 RT "Identification of a missense mutation in one allele of a patient  
 RT with Pompe disease, and use of endonuclease digestion of  
 RT PCR-amplified RNA to demonstrate lack of mRNA expression from the  
 RT second allele.";  
 RL Am. J. Hum. Genet. 49:635-645(1991).  
 RN [10]  
 RP VARIANT GSD-II LYS-521.  
 RX MEDLINE=91379015; PubMed=1898413;  
 RA Hermans M.M.P., de Graaff E., Kroos M.A., Wisselaar H.A., Oostra B.A.,  
 RA Reuser A.J.J.;  
 RT "Identification of a point mutation in the human lysosomal alpha-  
 RT glucosidase gene causing infantile glycogenosis type II.";  
 RL Biochem. Biophys. Res. Commun. 179:919-926(1991).  
 RN [11]  
 RP VARIANTS GSD-II ARG-643 AND TRP-725.  
 RX MEDLINE=94004908; PubMed=8401535;  
 RA Hermans M.M.P., Kroos M.A., de Graaff E., Oostra B.A., Reuser A.J.J.;  
 RT "Two mutations affecting the transport and maturation of lysosomal  
 RT alpha-glucosidase in an adult case of glycogen storage disease type  
 RT II.";  
 RL Hum. Mutat. 2:268-273(1993).  
 RN [12]  
 RP VARIANT GSD-II GLU-645, AND VARIANTS ILE-816 AND ILE-927.  
 RX MEDLINE=93168115; PubMed=8094613;  
 RA Hermans M.M.P., de Graaff E., Kroos M.A., Wisselaar H.A.,  
 RA Willemsen R., Oostra B.A., Reuser A.J.J.;  
 RT "The conservative substitution Asp-645-->Glu in lysosomal alpha-  
 RT glucosidase affects transport and phosphorylation of the enzyme in an  
 RT adult patient with glycogen-storage disease type II.";  
 RL Biochem. J. 289:687-693(1993).  
 RN [13]  
 RP VARIANT GSD-II GLU-645, AND VARIANTS ILE-816 AND ILE-927.  
 RX MEDLINE=92096118; PubMed=1684505;  
 RA Martinik F., Mehler M., Bodkin M., Tzall S., Hirschhorn K., Zhong N.,  
 RA Hirschhorn R.;  
 RT "Identification of a missense mutation in an adult-onset patient with  
 RT glycogenosis type II expressing only one allele.";  
 RL DNA Cell Biol. 10:681-687(1991).  
 RN [14]  
 RP VARIANTS ILE-816 AND ILE-927.  
 RX MEDLINE=93252406; PubMed=8486380;  
 RA Hermans M.M.P., Svetkey L.P., Oostra B.A., Chen Y.T., Reuser A.J.J.;  
 RT "The loss of a polymorphic glycosylation site caused by Thr-927-->Ile  
 RT is linked to a second polymorphic Val-816-->Ile substitution in  
 RT lysosomal alpha-glucosidase of American blacks.";  
 RL Genomics 16:300-301(1993).  
 RN [15]  
 RP VARIANT GSD-II VAL-519.  
 RX MEDLINE=95170739; PubMed=7866409;  
 RA Huie M.L., Hirschhorn R., Chen A.S., Martinik F., Zhong N.;  
 RT "Mutation at the catalytic site (M519V) in glycogen storage disease  
 RT type II (Pompe disease).";  
 RL Hum. Mutat. 4:291-293(1994).  
 RN [16]  
 RP VARIANT GSD-II TRP-647.  
 RX MEDLINE=95072571; PubMed=7981676;  
 RA Huie M.L., Chen A.S., Brooks S.S., Grix A., Hirschhorn R.;  
 RT "A de novo 13 nt deletion, a newly identified C647W missense mutation  
 RT and a deletion of exon 18 in infantile onset glycogen storage disease  
 RT type II (GSDII).";  
 RL Hum. Mol. Genet. 3:1081-1087(1994).  
 RN [17]  
 RP VARIANT GSD-II LEU-545.  
 RX MEDLINE=95187163; PubMed=7881422;  
 RA Hermans M.M.P., de Graaff E., Kroos M.A., Mohkamsing S., Eussen B.J.,  
 RA Joosse M., Willemsen R., Kleijer W.J., Oostra B.A., Reuser A.J.J.;  
 RT "The effect of a single base pair deletion (delta 7525) and a C1634T  
 RT missense mutation (Pro545Leu) on the expression of lysosomal alpha-  
 RT glucosidase in patients with glycogen storage disease type II.";  
 RL Hum. Mol. Genet. 3:2213-2218(1994).  
 RN [18]  
 RP VARIANTS GSD-II ARG-299; LYS-903 DEL AND VARIANTS HIS-199; ARG-223 AND  
 RP VAL-780.  
 RX MEDLINE=95233437; PubMed=7717400;  
 RA Boerkoel C.F., Exelbert R., Nicastri C., Nichols R.C., Miller F.W.,  
 RA Plotz P.H., Raben N.;  
 RT "Leaky splicing mutation in the acid maltase gene is associated with  
 RT delayed onset of glycogenosis type II.";  
 RL Am. J. Hum. Genet. 56:887-897(1995).  
 RN [19]

RP SEQUENCE OF 631-680 FROM N.A., AND VARIANT GSD-II HIS-645.  
 RX MEDLINE=95209708; PubMed=7695647;  
 RA Lin C.-Y., Shieh J.-J.;  
 RT "Identification of a de novo point mutation resulting in infantile  
 RL form of Pompe's disease.";  
 RN Biochem. Biophys. Res. Commun. 208:886-893 (1995).  
 RN [20]  
 RP VARIANT GSD-II VAL-529.  
 RX MEDLINE=96431168; PubMed=8834250;  
 RA Tsunoda H., Ohshima T., Tohyama J., Sasaki M., Sakuragawa N.,  
 RA Martinluk F.;  
 RT "Acid alpha-glucosidase deficiency: identification and expression of  
 RL a missense mutation (S529V) in a Japanese adult phenotype.";  
 RN Hum. Genet. 97:496-499 (1996).  
 RN [21]  
 RP VARIANTS GSD-II ASN-645, TRP-647; SER-648; GLN-672 AND TRP-672.  
 RX MEDLINE=98205825; PubMed=9535769;  
 RA Huie M.L., Teujino S., Brooks S., Engel A., Elias E., Bonthron D.T.,  
 RA Bessley C., Shanske S., Dinaiuro S., Goto Y.I., Hirschhorn R.;  
 RT "Glycogen storage disease type II: identification of four novel  
 RL missense mutations (D645N, G648S, R672W, R672Q) and two  
 RT insertions/deletions in the acid alpha-glucosidase locus of patients  
 RL of differing phenotype.";  
 RN Biochem. Biophys. Res. Commun. 244:921-927 (1998).  
 RN [22]  
 RP VARIANTS GSD-II PRO-566; ARG-643 AND ARG-768, AND VARIANTS ASN-91;  
 RP HIS-199 AND ARG-223.  
 RX MEDLINE=98180719; PubMed=9521422;  
 RA Hermans M.M.P., Kroos M.A., Smeitink J.A.M., van der Ploeg A.T.,  
 RA Kleijer W.J., Kruseur A.J.J.;  
 RT "Glycogen storage disease type II: genetic and biochemical analysis  
 RL of novel mutations in infantile patients from Turkish ancestry.";  
 RN Hum. Mutat. 11:209-215 (1998).  
 RN [23]  
 RP VARIANT GSD-II GLY-VAL-PRO-VAL-SER-ASN-925 INS.  
 RA Beesley C.E., Child A.H., Yacoub M.Y.;  
 RT "The identification of five novel mutations in the lysosomal acid  
 RL alpha-(1,4) glucosidase gene from patients with glycogen storage  
 RL disease type II.";  
 RN Hum. Mutat. 11:413-413 (1998).  
 RN [24]  
 RP VARIANT GSD-II ARG-481.  
 RX MEDLINE=99202470; PubMed=10189220;  
 RA Raben N., Lee E., Lee L., Hirschhorn R., Plotz P.H.;  
 RT Query Match 31.3%; Score 1492.5; DB 1; Length 952;  
 RT Best Local Similarity 37.8%; Pred. No. 1.1e-96;  
 RT Matches 339; Conservative 139; Mismatches 296; Indels 123; Gaps 26;  
 QY 40 YKVKSVKVDSTGRSITLALPOLVKNSSVYGP-DIQLLSITASLESNDRLRVRIIDAKHRR 98  
 DB 136 YKLENL---SSSEMGYTA--TLTRTPPTFFPKDILTLDVMMETENRHLFTIKDPANRR 190  
 QY 99 WEIPONILHRHPPPPPHSSSLYRLTLSSPTTNRKILLSHPNSDLTFSINTTPPGF 158  
 DB 191 YEVPLETRVHSRAPSPLYSVE-----FSEEPFGV 220  
 QY 159 TTSRSTHDVLPDAIPDTPNPNTLFLIDQYLHLTSSLPGRTHIYGLGHSKPTFQLAH 218  
 DB 221 IVHROLGDRVLLNTTVP-----LFFADQFQLSTSLPS--QYITGLAHLSPMLMST 272  
 QY 219 NOTLTMRAADIPSSPDVNLVGHSPFYMDVRSSPVAGSTHGVLLNSNGMDVEYTG-N-R 277  
 DB 273 WTRITLWNRDL-APTPGANLYGSHPPYLALED---GGSAHGVLFLNSNAMDVVLQSPAL 328  
 QY 278 TYKVIIGGIIDLYFFAGSPGQVVEQFTRVIGRPAPMPYAFQOCQRYGYHDVYELQSV 337  
 DB 329 SMRSTGILDVYIFLGPSPKVVQVYLDVGVYFPMFPYWGFLGFLHRCWCYSSTAITRQW 388  
 QY 338 AGYAKAKIPLVWNTDIDYNDAYKQFTLDPVNPFLDQMKPF---VNNLHNGCKTQVILD 394  
 DB 389 ENMTRAHFPLDQWDLMDYMSRRDPT-----FNKDGFRDFPQVQLHGGRRYTMIVD 443

## RESULT 7

AGLU SCHPO  
 ID \_AGLU SCHPO STANDARD; PRT; 969 AA.  
 AC Q9C0Y4;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase).  
 GN AGL OR SPAPB2403.10C.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A., SEQUENCE OF 25-36; 175-194; 375-395 AND  
 RP 427-451, AND MUTAGENESIS OF ASP-481; GLU-484 AND ASP-647.  
 RX MEDLINE=21195240; PubMed=11298744;  
 RA Okuyama M., Okuno A., Shimizu N., Mori H., Kimura A., Chiba S.;  
 RT "Carboxyl group of residue Asp647 as possible proton donor in  
 RT catalytic reaction of alpha-glucosidase from Schizosaccharomycetes  
 RT pombe.";  
 RL Eur. J. Biochem. 268:2270-2280 (2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwallian R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,

RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purrelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Gallbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerrutti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,  
RA Shpakowski G.V., Ussery D., Barrell B.G., Nurse P.,  
RT "The genome sequence of Schizosaccharomyces pombe.";  
RL Nature 415:871-880 (2002).  
CC -!- FUNCTION: Hydrolyzes malto-oligosaccharides, but has a low  
CC activity toward soluble starch.  
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-  
CC linked D-glucose residues with release of D-glucose.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC  
CC EXBL; AL590582; CAC35906.1; --  
CC EMBL; AB045751; BAB3946.1; --  
CC GeneDB Spombe; SPAB24D3.10c; --  
CC InterPro; IPR000322; Glyco hydro 31.  
CC Pfam; PF01055; Glyco hydro 31; 1  
CC PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1  
CC PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; FALSE\_NEG.  
KW Hydrolase; Glycosidase; Glycoprotein; Signal.  
FT SIGNAL 1 24  
FT CHAIN 25 969 ALPHA-GLUCOSIDASE.  
FT ACT\_SITE 481 481 NUCLEOPHILE.  
FT ACT\_SITE 484 484  
FT ACT\_SITE 647 647  
FT CARBOHYD 37 37 PROTON DONOR.  
FT CARBOHYD 67 67 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 99 99 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 139 139 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 146 146 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 245 245 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 249 249 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 331 331 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 406 406 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 429 429 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 462 462 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 470 470 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 520 520 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 523 523 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 529 529 O-LINKED (POTENTIAL).  
FT CARBOHYD 543 543 O-LINKED (POTENTIAL).  
FT CARBOHYD 589 589 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 648 648 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 801 801 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 810 810 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 821 821 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 885 885 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 915 915 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 934 934 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 942 942 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 954 954 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 966 966 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CONFLICT 30 30 L -> F (IN REF. 2).  
FT CONFLICT 220 220 P -> A (IN REF. 1).  
FT CONFLICT 507 507 T -> V (IN REF. 1).  
FT CONFLICT 566 566 D -> N (IN REF. 1).  
SQ SEQUENCE 969 AA; 108686 MW; F3122E2CFA551C25 CRC64;  
Query Match 30.7%; Score 1464.5; DB 1; Length 969;  
Best Local Similarity 35.8%; Pred. No. 1e-94;  
Matches 335; Conservative 157; Mismatches 284; Indels 159; Gaps 27;  
QY 31 DP-EGVIGYGVKVKVGVKSGTGRSLTA-LPOLVKNSSVYGPDIQLLSITASLESNDRLR 88  
DB 53 DPRGVNCOGQAVNI---SESONGTAVLALLGCEPCYAGTDYPLFLFNVTYEADRVH 109  
QY 89 VRITDAKHRWEIPDNILHRHQPPPPPHSLXRTLLSSPTTNRKILLSHPNSDLTF 148  
DB 110 ISIKDANNTQFQT-----SRKLDWAPLYSPSYNNNTLLYN-----F 147  
QY 149 SLINTTPEGFTISRKSTHVDLFDATPDPTNPNTFLIFIDQYLHLTSSLPTCAHIYGLGE 208  
DB 148 S-YNAMPFEFWTRKSDGEVLFDTRQK-----LVFEDQYIELTTNMV-ENVNLVGLAE 199  
QY 209 --HSKPTFQLAHNQTLTMRADIPSSNPDPVNLVYGHSPFYMDVR-----SSPVAG 255  
DB 200 TIHG---LRLGNLTLTFWANDPSP-VQONMYGSHPYLQRYKADGINSLTNETTYS 255  
QY 256 STHGVLILNSGMVDYETGNRTYKIVGIIIDYFFAG--PSPGVVEQFTRVIGRPAPM 313  
DB 256 SSHGVMLTANGDVLRLDQYLMIGVIGVILFVYSGSTESPKETVKQFVOSIGKPAH 315  
QY 314 PYAFGQOCRCYCHVDYVYELQSVYACVAKIPLVMTDIDYMDAYKQFTLDVNVFPLD 373  
DB 316 QYMTLGHSCRWGTYTITEIMDVQRQYIDADIPVEITFMSDIDIMEKIRFTVDVPVSIKS 375  
QY 374 KMKKFNVLNHNKQKQVVLDPGI-----STNKTYTYIRGMKHDFLKR-NGPKYLG 425  
DB 376 DMQTFESDLVSNHGVYPIIDAALYAANPVNHTDDSYYPYAGVEKDIFLKNPNSIYG 435  
QY 426 SVWPGVPYFDPFLKPSALFTWDEIKRFLNL-----LPVDGLWIDMNEISNF---- 472  
DB 436 AVWPGTAFDPFTNPVDVYWKDCL---INLTYAFGNSGTVPFSGIWTDMNEPSSPCVGS 492  
QY 473 -----ISSP 476  
DB 493 CGSAMIDLNPABPLTGISKOYSIPEGFVNSVNTVEYSASLSNYYATATSSVFQIVSP 552  
QY 477 ---PI---PGSTLDNPPYKINNSGVMLPIINKTPTTAMHYGDIPEYNVHNLEGYLEARY 530  
DB 553 TATPLGLKPDYNIWDPPYAINNEQNHDIANHIVSPNATTHDGTQRYDIFNMYGYGETKV 612  
QY 531 TRAAIKLT-EKPPFVLSRSTSGSKYTAHWTDGNAATWDLVSIIPSMDFGLGIPM 599  
DB 613 SYAALTQISPNERPFILSRSTFLGSGYGAHWLGDHSLNSMFFSISGMIVFMMGIPM 672  
QY 590 VGADICQFLGNTTEELCRRWIQLCAFPSPRDHSSLGTTTYQELYRWESVAASARKVLGLR 649  
DB 673 VGADVCGFLGDSBELCSRMMANGAPSPFRNHNIIYQISQEPYTWSSVAEASPRAMYIR 732  
QY 650 YTLPPYFTLMTYEAQLNGIPIARPLFFSPDDIKTYGISQFLGKGVMVSPVLKPGVWS 709  
DB 733 YSLLPYWTIMAKASQDGTPLRALPVEFPDPTLADVDQRQFMVGDLSLLVTPLEPNVEY 792  
QY 710 VTAYFPRGN---WFDLFDTYRSTASTGRYVTLSPDPHNVHIQEGNILLAMQKAWTTQ 766  
DB 793 VQGVFPGDNSTVMYDWNHT-EIVRQYNNVNTYAPLEHINVAIRGSGVLPMPQPSLTYY 851  
QY 767 AAKTPTPHLLVMSDCGASPGELFDGGEVTVGVNRGKWTFFVKFIAAASAKQTCITSDV 826  
DB 852 ESKQNPFLVALDRGSGATGELYLDGSGIELNATLS-----VSFTFSDG-----VLSAV 902  
QY 827 VSGEFAVSQKWIDKVTILGRKGTG---INGYTV 858

903 FTGSYVQSP--LANVTILGLTESPSSITLQGNV 935

Db 903 FTGSYVQSP--LANVTILGLTESPSSITLQGNV 935

RESULT 8

ANYG\_CANAL STANDARD; PRT: 946 AA.

AC 074254;

DT 15-JUL-1999 (Rel. 38, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Glucoamylase 1 precursor (EC 3.2.1.3) (Glucon 1,4-alpha-glucosidase)

DE (1,4-alpha-D-glucon glucohydrolase).

GN GAM1 OR GCA1.

OS Candida albicans (Yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales, mitosporic Saccharomycetales; Candida.

OX NCBI\_TaxID=5476;

RN [1]

SEQUENCE FROM N.A.

RP STRAIN=SC5314;

RX MEDLINE=99451422; PubMed=10520161;

RA Sturtevant J., Dixon F., Madsen E., Latge J.-P., Zhao X.-J., Calderone R.;

RT "Identification and cloning of GCA1, a gene that encodes a cell surface glucoamylase from Candida albicans.";

RL Med. Mycol. 37:357-366(1999).

CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-glucose residues successively from non-reducing ends of the chains with release of beta-D-glucose.

CC -!- SUBCELLULAR LOCATION: CELL WALL ASSOCIATED.

CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch>) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC -----

DR EMBL: AF082188; AAC31968.1; "

DR InterPro: IPR000322; Glyco\_hydro\_31.

DR Pfam: PF01055; Glyco\_hydro\_31; 1.

DR PROSITE: PS00129; GLYCOSYL HYDROL\_F31\_1; 1.

DR PROSITE: PS00707; GLYCOSYL HYDROL\_F31\_2; 1.

KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein; Signal; Cell wall.

FT SIGNAL 1 20 POTENTIAL.

FT CHAIN 21 946 GLUCOAMYLASE 1.

FT ACT\_SITE 462 462 BY SIMILARITY.

FT DOMAIN 519 532 SER/THR-RICH.

FT CARBOHYD 51 51 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 97 97 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 244 244 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 373 373 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 393 393 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 406 406 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 437 437 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 505 505 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 570 570 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 772 772 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 801 801 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 895 895 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 912 912 N-LINKED (GLCNAC. .) (POTENTIAL).

SQ SEQUENCE 946 AA; 105804 MW; BD6B640CEE1F70 CRC64;

Query Match 30.0%; Score 1431.5; DB 1; Length 946;

Best Local Similarity 34.9%; Pred. No. 2e-92;

Matches 335; Conservative 150; Mismatches 310; Indels 165; Gaps 30;

QY

Db 32 OKTVPVELSIG-----VKQLPNHNSAVDANAVAKGYSLVNSL-----TARGLTG 78

QY 58 LPQLVKNSSVYGPDIQLLSITASLESNDRLVRITDAKHRRWEIPDNILHRHQPPIPPH 117

Db 79 ILKLEANTNIYGYDFEYLNLSVEYQSDTLNVIH-----EPT 115

QY 118 SISSLY---RTLSSPTNRRKILLSHPNSDLTFSLINTPPGFTISRKSTHDVLFDATP 174

Db 116 DLTDVFLPEELVVPKLEGDAKTFNFENSDLVFE-YDEEDGFVFLRSSREVLFPSTKG 174

QY 175 DPTNPNTFIIDQYLHLTSSLPGTRAH-IYGLGSH-----SKPTFQLAHNQTLMRAAD 228

Db 175 NP-----LVFSNQFIQFTLTP--KGHSITGLGESIHGSLNEP-----GVVKLYAND 220

QY 229 IPSSNP-DVNLGSGHPFYMDVRESSPVAGSTHGVLNLSNGMDVEYTGNRITTKYVGGIID 287

Db 221 I--ADPIDGNIYGVHPVYDQRYD--TNTHGVWRTSAIQEVVVGSETSLTWRALSGVID 276

QY 288 LYFFAGSPGQVVEQFTRVIGRPAPMPYMAFGQCRGYHDVYELQSVVAGYAKAKIPL 347

Db 277 LYFFSGDPDKVIQYVSEIGLPAMQPYWALGYHOCRWGYDTVESLETVENFKKFDIPL 336

QY 348 EVMWTDIDVMDAYKOFITLDPVNPFLDKMKKFWNLHKGQKIVVILD-----PGISTN 400

Db 337 ETIWSIDIDYMGYKOFITNDPYRPTDKFRKFLDDLHNSQHYVPIFDAAIYVVPNNATD 396

QY 401 KYVETVIRGKMDVFLKR-NGKPYLGSVWPGVPYFPDFLKPALSALTFTWDEIKRFLNLLPV 459

Db 397 DYEPFLHGNESDVLKNDPGSLYIGAVWPGYTVFPDFLANNTQYWNQKFWKVERIFP 456

QY 460 DGLWIDMNEISNP-ISS-----PPI-----478

Db 457 DGIWTDMEVSSFCVSGSCGTDRYDFNPVHPPEVGYSGSDYPLGDKSNASEWKSISEAA 516

QY 479 -----PG-STLDNPPYKINNSGYMLPIINKTIPTTAHYCD 513

Db 517 AATKTTTTSSASTSIDGKNTLAPGKGNINYPYAINNDQGDHDLATHAISPNATHADG 576

QY 514 IPEYNVHNLFGYLEARVTRAALKL-TEKRPVLSRSTFGSGSKYTAHWTGDAATWNDL 572

Db 577 TVEYDHNIIYGLIERAIYEALLEHPKRPFIIGRSSFAGSKYMGWGGDNYADYNNM 636

QY 573 VYSIPSMDFLFGIPMVGADICGLGNTEELCRRWIOLGAFYPPSRDHSLLGTYQEL 632

Db 637 YFSIPQALSMGLSGIPFGVDACGNGNTMELCSRWMLASFPFPRNHNVLGAIPQEP 696

QY 633 YRWESVAASARKVLGLRYTLTLLPYTLMTYAOLNGIPIARPLFFSPDDIKTYGSSQPL 692

Db 697 YWWEAYMKAITKSINVRYSLLPYTLHSHVGTGIPINRAFNWQFPYSKELAGVDTOFF 756

QY 693 LGKGVNVSPVLKPGVSVTAYFPRGN--WFDLFDVTRSVTASTGRYVTLISAPDPHINVI 750

Db 757 VGDALLVTPVLEPGVNVHTKGIFFGENAVYDYTHKKQ-KFTAGKNETLAAPLGHILPHI 815

QY 751 QEGNILLAMQKAWTTOAARKTFFHLLVMSDCGASFGELFDGGEVTEVMGNVRGKWTYK 810

Db 816 KGGNIPTQEPGYTTTESRKNPFGLIIVALDASGTASGLYLDGGSVDV---EEALYVD 871

QY 811 FTAASAKOTCIITSDVSGEFAVSQKWDKVTILGLRKGTK---INGYTV---RTGAV 863

Db 872 FVASKNK---LVAS--VFGEYVRQP--LANVTILGVDSBPKKVLFPNNETSHKYGAV 924

RESULT 9

AGLU\_ASPOR STANDARD; PRT: 985 AA.

ID\_AGLU\_ASPOR

AC Q12558;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase) (AGL).

GN AGDA.



RA Masaki H., Kimura A., Chiba S., Uozumi T.;  
 RT "Cloning and sequencing of an alpha-glucosidase gene from Aspergillus  
 RL niger and its expression in A. nidulans";  
 RN J. Biotechnol. 53:75-84 (1997).  
 RP [2]  
 RC SEQUENCE OF 26-252 AND 267-985.  
 RX STRAIN=GN-8;  
 RA MEDLINE=93005089; PubMed=1368849;  
 RA Kimura A., Takata M., Sakai O., Matsui H., Takai N., Takayanagi T.,  
 RA Nishimura I., Uozumi T., Chiba S.;  
 RT "Complete amino acid sequence of crystalline alpha-glucosidase from  
 RL Aspergillus niger";  
 RL Biochem. Biotechnol. Biochem. 56:1368-1370 (1992).  
 CC -!- FUNCTION: Hydrolyzes malto-oligosaccharides, but has a low  
 CC activity toward soluble starch.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-  
 CC linked D-glucose residues with release of D-glucose.  
 CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC  
 CC EMBL; D45356; BAA23616.1; -;  
 DR GlycoSuiteDB; P56526; -;  
 DR InterPro; IPR000322; Glyco\_hydro\_31.  
 DR Pfam; PF01055; Glyco\_hydro\_31; 1.  
 DR PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
 DR PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.  
 KW Hydrolase; Glycosidase; Glycoprotein; Signal.  
 FT SIGNAL 1 25  
 FT CHAIN 26 985 ALPHA-GLUCOSIDASE.  
 FT ACT\_SITE 490 490  
 FT CARBOHYD 36 36 O-LINKED (POTENTIAL).  
 FT CARBOHYD 124 124 N-LINKED (GLCNAC. ).  
 FT CARBOHYD 143 143 N-LINKED (GLCNAC. ).  
 FT CARBOHYD 218 218 N-LINKED (GLCNAC. ).  
 FT CARBOHYD 347 347 N-LINKED (GLCNAC. ).  
 FT CARBOHYD 422 422 N-LINKED (GLCNAC. ).  
 FT CARBOHYD 506 506 N-LINKED (GLCNAC. ).  
 FT CARBOHYD 534 534 N-LINKED (GLCNAC. ).  
 FT CARBOHYD 537 537 N-LINKED (GLCNAC. ).  
 FT CARBOHYD 545 545 O-LINKED (POTENTIAL).  
 FT CARBOHYD 550 550 O-LINKED (POTENTIAL).  
 FT CARBOHYD 559 559 O-LINKED (POTENTIAL).  
 FT CARBOHYD 560 560 O-LINKED (POTENTIAL).  
 FT CARBOHYD 561 561 O-LINKED (POTENTIAL).  
 FT CARBOHYD 562 562 O-LINKED (POTENTIAL).  
 FT CARBOHYD 601 601 N-LINKED (GLCNAC. ).  
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. ).  
 FT CARBOHYD 835 835 N-LINKED (GLCNAC. ).  
 FT CARBOHYD 881 881 N-LINKED (GLCNAC. ).  
 FT CARBOHYD 899 899 N-LINKED (GLCNAC. ).  
 FT CARBOHYD 957 957 N-LINKED (GLCNAC. ).  
 FT CARBOHYD 970 970 N-LINKED (GLCNAC. ).  
 FT VARIANT 27 28 TT -> LL (IN STRAIN GN-8).  
 FT VARIANT 42 42 D -> A (IN STRAIN GN-8).  
 FT VARIANT 929 929 N -> M (IN STRAIN GN-8).  
 SQ SEQUENCE 985 AA; 108913 MW; 9A18772AEB2E0927 CRC64;

Query Match 29.4%; Score 1401.5; DB 1; Length 985;  
 Best Local Similarity 33.4%; Pred. No. 2.7e-90;  
 Matches 327; Conservative 160; Mismatches 287; Indels 205; Gaps 29;  
 QY 21 LVAGYVSTENDPEGV----ICGYKXVKSVKVDGSRFRSLTALPOLV-KNSVWGPDQLL 75  
 DB 47 LIANI-----DDQADAQSVCEGYKASKVQNS---RGFTASLQAGPCVYIGDVESL 99  
 QY 76 SITASLESNDRLVRIRI----TDAKHRRWE-IPDNILHRHQPPPPPPPSLSLRYLTLLSSP 130

DB 100 TLISVQSDRLNIQILPHVDSTNASWVPLSENLV-----PRKASINA----- 144  
 QY 131 TTRNRKILLSHPNSDLTSLINTPTPGFTIRKSTHDLVLPATPDPPTNPFLIFIDQVL 190  
 DB 145 -----SVSQSDFVSWSNPSFNFVKIRKATGDAIF-----STEGTVLVYENOFI 189  
 QY 191 HLTSSLPGTGRAHYGLGHSKPTFOLAHNOTLTWRAADI PSSNP-DVNLVYSHSPYMD-- 247  
 DB 190 EFWTALP-BEYMLYGLGEHI-TQFRLQRNANLTIYPSD-DGTPIDQNLVGOHPYILTR 245  
 QY 248 -----VRSSPVAGS-----THGYLLNSNGMDVEYTGNRITYKVGIGIIDL 288  
 DB 246 YYKGRQNGSYIPVKSSEADASQDYSILSHGVFLNSHGLEILLRSQKLIWRTLGGIDL 305  
 QY 289 YFAGSPSQVVEQF-TRVIGRPAMPYWAGFOQCRYGYHDVYELQSVWAGYAKAKIPL 347  
 DB 306 TFSGPAPADVTRQYLTSTVGLPANQQYNTLGFHOCRWGYNMWSLDADVANFEKFIPL 365  
 QY 348 EVMWTDIDYMDAYKDTLDPVNPFLDKMKKFWNNLHKQOKYVWILDPIG-----STNK 401  
 DB 366 EYIWTDDYMHGTRNPDNDQHRFSYSEGDFLSKLHESGRYVYVPIVDAALYIPNENASD 425  
 QY 402 TYETVIRGMKHDFVLR-NGKPYLGSVMPGPVYVPPDFLKPASLTFTWDEIKRFLNLLPD 460  
 DB 426 AYATYDRGAADVFLKNPDGSLYICAVMPGYTVFPDHPHKAVDFWANELVIWKKVAFD 485  
 QY 461 GLWIDMNEISNF----- 472  
 DB 486 GWYDMSEVSFCVSGCGTGNLTNPAHPSFLLPGEQDIIYDPEAFNITWATEAASAS 545  
 QY 473 -----ISSPPIPG-STLDNPPYKINNSGVMLPIINKTIPPTAMHY 511  
 DB 546 AGASSQAARATTTSTSVSYLRTTTPGVRNVHPYVYINHDEQGHDLNVHAVSPNATHV 605  
 QY 512 GDIPENYNNLFGYLEARVTRAALIKL--TEKRPVFLSRSPFSGSGKYTAHWGTGNAATV 569  
 DB 606 DGVEEYDVHGLYGHQGLNATYQGLLEVWMSHKRRPFIIGRSTFAGSGKWAGHWGDNYSKW 665  
 QY 570 NDLVYSIPSLMDPLFGIPMGADICGLNGNTEELCRRWIOLGAFYFSPRSHSLSLTGY 629  
 DB 666 WMSYYSISQALSFSLFGIPMGADTCGFNGNDEELCRNMOLSAFFFFPYRNHNLSTIP 725  
 QY 630 QELYRWESVAASARKYGLRYTLPLFYFTLYEAQINGIPIARPLFFSPDPDIKTYGSS 689  
 DB 726 QEPYRWASVIEATKSAMRIRYAILPFYFTLFLDHAHTTGSTVNRALSWEFPNDTLAAVET 785  
 QY 690 QFLLGKGVWSPVLKPGVSVSTAYFP---RGN-WFDLFDYTR-SVTASTGRVYVTLSPDP 744  
 DB 786 QFMVGPAMVVPVLEPLVNTVKGVPFGVGHGEVWYDW--YTQAAVDAKPGVNTTISAPLG 843  
 QY 745 HINVHIQEGNIIAMOCKAMTTOARKTPFHLLVMSDCGASFGELFDDGVEV---TWG 800  
 DB 844 HIPVTVRGNILLPMQEPALUTTREARKTPWALLAALNGSNGTAGSQLYLDDGESIYPNATHL 903  
 QY 801 VN-----RGKTVFVKFIAASAKQTCIITSDVVSGBFAVSQKVIDKVTILGLR 848  
 DB 904 VDTASRSLSRSAQGRW-----KERNPLANVTVLGVN 936  
 QY 849 K---GTKINGYTVRGAVT 864  
 DB 937 KEPSAVTLNGQAVFPGSVT 955  
 RESULT 11  
 YAJI\_SCHPO  
 ID YAJI\_SCHPO STANDARD; PRT; 993 AA.  
 AC Q09901;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Putative family 31 Glucosidase C30D11.01c precursor (EC 3.2.1.-).  
 GN SPAC30D11.01C OR SPAC56F8.01.

OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;

RA Wood V., Gwilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitz E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,  
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas R., Rochet M., Gallard J.C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Senito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Useery D., Barrell B.G., Nurse P.;  
 RA "The genome sequence of Schizosaccharomyces pombe";  
 RL Nature 415:871-880(2002).

CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; Z67961; CAA91887.1; -

DR EMBL; Z69728; CAA93572.1; -

DR PIR; T38598; T38598.

DR GeneDB Spombe; SPAC30D11.01c; -

DR InterPro; IPR00322; Glyco hydro 31.

DR Pfam; PF01055; Glyco hydro 31; 1

DR PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; FALSE\_NEG.

DR PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.

KW Hypothetical protein; Hydrolase; Glycosidase; Glycoprotein; Signal.

FT SIGNAL 1 24 POTENTIAL.

FT CHAIN 25 993 PUTATIVE FAMILY 31 GLUCOSIDASE

FT DOMAIN 24 36 POLY-SER.

FT CARBOHYD 7 7 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 44 44 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 89 89 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 121 121 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 138 138 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 161 161 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 169 169 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 232 232 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 361 361 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 386 386 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 393 393 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 447 447 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 480 480 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 545 545 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 548 548 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 614 614 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 673 673 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 814 814 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 826 826 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 835 835 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 846 846 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 910 910 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 940 940 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 987 987 N-LINKED (GLCNAC. .) (POTENTIAL).

SQ SEQUENCE 993 AA; 111043 MW; EB9471F19AAA9BC9 CRC64;

Query Match 28.9%; Score 1379; DB 1; Length 993;

Best Local Similarity 33.3%; Pred. No. 1e-88;

Matches 312; Conservative 170; Mismatches 287; Indels 168; Gaps 25;

Qy 22 VAGISTSENDPEGVIGYGVKSVKSDTRSLTALPOLVRNSS-VYGPDIQLLSITAS 80

Db 67 VAQETLYESSRGLSCPGYOARNI---SEYGVLAILELAGDACVAYGTDYPLLNV 123

Qy 81 LESNDRLVRITDAKRRWEIPDNILHRHQPPPPPHSLSSLYRTLSSPTTNRKIL-- 138

Db 124 YDTEERVHISISDLNQTQFL-----SNRRDWDWA 153

Qy 139 -LSHPNSDLTFSL-----INTTPFGFTISRKSTHVDLFDATPPTNPNTFLIPDOYLHL 192

Db 154 PLFYRSNFGNLOYNFSFNTDPPFWITRIADDQVLFTRGNP-----LIFEDQYIEL 207

Qy 193 TSSLPOTRAHIVGLGBHSHKPTQLAHNQTLRADIPSNPDVNLVGHSPYMOVRSPP 252

Db 208 TTNMV-EDYNYVGL-SGSQSQFRLGNLTKTWATGY-SDSPEANNYGHSPMEQRYIP 264

Qy 253 V-----AGSTHGVLNLSNGMDVEYGNRTYKVIGGIIDLYFFAGP--SPQVVEQFT 304

Db 265 IGTNTYTSASHGVMLSSNGMEVLRSYIKYRMIGGIIDLVFVSGSVSPRYTITQQV 324

Qy 305 RVIGRPAPMPYAFGQOCRYGYHVDVYELQSVVAGYAKAKIPLVEMWTIDYMDAYKDT 364

Db 325 QSIGTPTMPQYWSLGFQMSRWGYKTLSDLINRMSYLNASNIPTGFWNDIDYSEFRFTT 384

Qy 365 LDPNVFPLDKMKHFNANLHNKNGQYVILDPGI-----STNKTYYETIRGMKHVFLK 417

Db 385 VNSTAPPNQTLDFFRSLDESHQHYVPVLDPAIYAANPNKSAADRTYYPYSGEDNIFIK 444

Qy 418 R-NGKPYLGSVWEGPYVPDFELKPSALTFTWTDIKFLNL-----LPVGLMID 465

Db 445 NPNGSAYGVGMAWPGFVYVPDFTNPVLOQWKGQI---LNLSTAFGNSYVDLFFSGLCLD 501

Qy 466 MNEISNF-ISS-----PP--IPGS-----481

Db 502 MNEPTSFICSGCDLLKLNVPVHPFSLPGDVNDKVSYPEDFNATNTEYKSVSRASQS 561

Qy 482 -----TLDNPPYKLNNGVMLPIINKTIPPTAMHYGDI 514

Db 562 QYKATATSEKSHETPSSESLINGKPEFSINYPYALDITDTETDLAQGVSPNATHGNT 621

Qy 515 PEYNVNLPGYLEARVTRAALIKLTK-RPFLVLSRSTFSGSGKYTAHTGDNAAATNDLV 573

Db 622 LRNLFTYGYSEKISFEALNIQNIPIRFLRSRTFVSGRYAAHAWLGNKSKQSDMW 681

Qy 574 YSIPSMILDFLFGIPMVGADICGFLGNTTEELCRRIQLGAFYPPSRDSSLGTTTYQELY 633

Db 682 SSISILITNLGI PMVGDVCGYNGTDEELCARWALGAFLPFYRNHNSLGSIPQEPF 741

Qy 634 RWSVAASARKVGLRYTLPLPYETLMYEAQLNGIPIARPLFFSPDDIKTYGISQFL 693

Db 742 RWSVAASARSATIRYSLPYETLMYEAQLNGIPIARPLFFSPDDIKTYGISQFL 802

Qy 694 GKGVMYSPVLPKGVSVTVAYFRGN---WFDLFDYTRSVTASGRVYVTLSPADHNVHI 750

Db 802 GTALLISPALEPNNTYIQGIIPGNDNTIMYDWNHS-VINHIDENITMSAPLYGVNI 860





Db 175 EYQTSNRFHFKLTDTQNNREVP-----HEH-----VQSF----- 205

Qy 140 SHPNSDITFSL-INTTFGFTISKSTHDLVFDATPDTPNTFLIPIDQYLHLTSLPG 198

Db 206 GNAASLTQVEISQRFPSIKVTRSRNNRVLDFSSIGF-----LLPADQFLQLSTRLPS 259

Qy 199 TRAHYGLGHSKTFQFOLAHN-QTLTWRAADIPSSNPDVNLGSHPPYMDVRSPVAGST 257

Db 260 T--NVYGLGEHVHQYRHDMMWKTWPIFNREDTPNGNGTNYLGAQTFFLCLEDA--SGLS 315

Qy 258 HGVLLNSNGMDVE-YTCNRITYVIGIGIIDLFFAGSPQCVQVEQFTRVIGSPAPMPY 316

Db 316 FGVFLMNSAYEVVLQAPALTYRTIGLLDFYVFLGNTPEQVQVEYLEJIGRPAUPSYW 375

Qy 317 AFGQOCRYGHDVYELQSVVAGYAKAPILEVWMTDIDYMDAYKDTLPDPVPPDLNKM 376

Db 376 ALGFHLRYEYGTLDNNREVVNRRAAQLPDVQADIDYMDERRDFTYDSVDF--KGFP 433

Qy 377 KFNANLHNGOKYVVLDPG-STN-----KTYETVIRGMKHDVFLKRG--KPYLGSWP2G 430

Db 434 EFVNELNHNGOKLVIIIDVPAISNNSSSKPYGPDYRGSMDKMWVSSDGVTPLIGVWPG 493

Qy 431 PVYFPDFLKPALFTWDEIKRFLNLLPVDGLMIDMNEISFISSPPIPGST--LDNPPY 498

Db 494 QTVFPDYTPNCAVMWTKPELPHNQVEFGDGMIDMNEVSNFVDSVSGCGSTNNLANNPF 553

Qy 489 KINNSGWLPP-----IINKTIPPTAM-HYGDIDPEYVWNLFGYLEART-RAALIKLTE 540

Db 554 -----TPRLDGYLFTCLCDVQHWG--KQYDIHNLGYSMVATAEAATVFPN 603

Qy 541 KRPVLSRSTFSGSKYTAHTGDNAAWDLVYSIPSMDFGLFGIPYMGADFCGFLGN 606

Db 604 KRSFILTRSTFAGSGKFAAHLWLGNTATWDLRWSIFQVLEFNLFGLPMWGPDCGFPALD 663

Qy 601 TTEELCRMIQLGAFYPSRDSHSLGITYOELRYWESVAA-----SARKVLGLJRYT 651

Db 664 TPEELCRMMQLGAFYPSRHNHGG-----YKQDPASFGADSLNLSRHYLNIRYT 717

Qy 652 LLPFYTLMAEQINGPIAPLFFSPDDIKTYGISQFLLGGMVSPVLKGVSVVT 711

Db 718 LLPYLYTLFFRAHSGDTPVARPLHFEFYEDNSTWDVHQQFLWGFGLLITVPLDGAEKVM 777

Qy 712 AYFPRGAWDFDYTRSVTASTGRY---VTLSPAPDHINVHIQEGNLAQGMKMTQA 767

Db 778 AYVDPANWYD-----ETGQVRRKQKVEMLPGDKIGLHKGVIFFTPQPNITILA 831

Qy 768 ARKTPFHLVVMSPCGASFGELFDGGEVTVTMVGNRGKWFVKFIAASAKQTCIITSVV 827

Db 832 SRKNPLGLIILADENKEAKGELFWDGGE-----TKDVT 864

Qy 828 SG-----EFAVSO-----KWVIDKVTILGRKGTCKINGVTVTGAVTR 865

Db 865 ANKVYLLCEFSYQNRLEVNISQSTYKDPNNLAFNEIKILGTEEPSNV-----TVKH 916

Qy 866 KGDGSKLSTPDRKGEFVAEISGLNLLGREF 898

Db 917 NGVFSQTSPTVYDSNLKVAITDIDLLGEAY 949

RESULT 13

AMYG DEBOC  
 ID AMYG DEBOC  
 AC P22861; 092336;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-AUG-1991 (Rel. 19, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Glucoamylase 1 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)  
 DE (1,4-alpha-D-glucan glucohydrolase).  
 GN GAMI.  
 OS Debaryomyces occidentalis (Yeast) (Schwanniomyces occidentalis).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Debaryomyces.

Query Match 28.8%; Score 1374; DB 1; Length 958;  
 Best Local Similarity 33.8%; Pred. No. 2.2e-88;  
 Matches 328; Conservative 142; Mismatches 315; Indels 184; Gaps 27;

Qy 3 KKIPSLALGILLVFLLOYL-----VAGISSENDPEGVI----- 36  
 Db 5 KLIKSVIGLGLVSAIQAAPASSIGSSASSASSSQATIPNDVTLGVKQIPNIFNSA 64  
 Qy 37 -----GYGVKYSKVDGSTRSLTALPOLVKNSSVYGPDIQLLSITASLESNDRLVR 90  
 Db 65 VDANAAGGYDLVNV---TNTPRGLTKLQKATNIYGYDFDYLNLTVGYQAPTRLNVH 121  
 Qy 91 ITDAKHRRWEIPDNLHRHQPPPPPHSLSSLY---RTLSSPTTNRKILLSHSPNSDLT 147

NCBI\_TaxID=27300;  
 [1] SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN=ATCC 26076;  
 RX MEDLINE=91071592; PubMed=1979298;  
 RA Dohmen R.J., Strasser A.W.M., Dahlems U.M., Hollenberg C.P.;  
 RT "Cloning of the Schwanniomyces occidentalis glucoamylase gene (GAMI)  
 RL and its expression in Saccharomyces cerevisiae.";  
 RN Gene 95:111-121(1990).  
 RP SIMILARITY TO OTHER FAMILY 31 MEMBERS.  
 RX MEDLINE=92077121; PubMed=1743281;  
 RA Naim H.Y., Niermann T., Kleinans U., Hollenberg C.P.,  
 RA Strasser A.W.M.;  
 RT "Striking structural and functional similarities suggest that  
 RT intestinal sucrase-isomaltase, human lysosomal alpha-glucosidase and  
 RT Schwanniomyces occidentalis glucoamylase are derived from a common  
 RT ancestral gene.";  
 RL FEBS Lett. 294:109-112(1991).  
 CC -!- FUNCTION: THIS GLUCOAMYLASE HAS A SPECIFICITY TOWARD BOTH  
 CC ALPHA-1,4 AND ALPHA-1,6 LINKAGES  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-  
 CC glucose residues successively from non-reducing ends of the chains  
 CC with release of beta-D-glucose.  
 CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; M60207; AAA33923.1; -  
 DR FIR; JN0102; JN0102.  
 DR InterPro; IPR000322; Glyco\_hydro\_31.  
 DR Pfam; PF01055; Glyco\_hydro\_31; 1.  
 DR PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
 DR PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.  
 KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;  
 KW Signal.  
 FT SIGNAL 1 22 POTENTIAL.  
 FT CHAIN 23 958 GLUCOAMYLASE 1.  
 FT ACT\_SITE 470 470 BY SIMILARITY.  
 FT DOMAIN 26 41 SER-RICH.  
 FT DOMAIN 530 542 SER/THR-RICH.  
 FT CARBOHYD 61 61 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 78 78 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 107 107 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 197 197 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 403 403 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 416 416 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 513 513 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 580 580 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 602 602 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 813 813 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 907 907 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 958 AA; 106507 MW; 47938DB9BC308260 CRC64;

```
Db 122 I-----EPTDLSOVFLPEHLVVKPLVEGDAQSYNFKRSDLV 158
Qy 148 FSLINTTFGFTISRKSTHVDVLPATDPDTPNPNTFLIFIDQYHLTSSLPCTRHYVGLG 207
Db 159 FEYSN-TDFSEVIRSTKEVLFSTKGNP-----LVFSNQIFQNSLSKNHV-I7GLG 210
Qy 208 EH-----SKPTFQLAHQNTLTMRAADTPSSNP-DVNLYGSGHPFMDVRSSPVGASTHGV 261
Db 211 ESIHGLVNEP-----GSVKTLFANDV--GDPIDGNIYGVHPVLDQRYD--TETHAVY 260
Qy 262 LLNSGMDVETGNRIYTKVIGGIDLYFPAGSPGQWQFTRVIGRPAPMPYAPGFQ 321
Db 261 WRTSAIOEVLIGESITWRALSGVIDLYFFSGPTPKDAIQOYVKEIGLPAFPQYMSLGYH 320
Qy 322 QCRGYHDVVELQSVAGYAKAKIPLEVMWTDIDYMDAYKDFLDVNEPFLDKMKKPVAN 381
Db 321 QCRGYDTIEKLSEVENFKFNPLETMSDIDYMDSYKDFDYDHRFPLDEYRKLDE 380
Qy 382 LHKNGQKVVILD-----PGISTNKTYYIRGMKHDVFLKR--NGKPYLGSVMPGPVY 433
Db 381 LHKNQHVPLDAAIYVPPNNAATDNEYQPFHYGNETDVLKPNDCGLYIGAWQ--VTL 439
Qy 434 PPDEL--KPSALTFTWDEIKRFLNLLPVDGLWIDMNEISNP-----ISPPIP-- 479
Db 440 FSRLSRKHSOM--DKVIKDWYBLTPFDGIWADMNEVSSFCVSGCGTGKYFENPAPYPPF 496
Qy 480 -----GSTILD----- 484
Db 497 TVGSKATSPYGVFDVSNASEWKSIOSSISATAKTSSSSSSSTTDYNTLAPGKNI 556
Qy 485 -NPYKINNKGMLPIINKTIPPTAMHYGDIPEYVNHNLFGYLBARVTRAALIKL-TEKR 542
Db 557 NYPVYAIYMQGSDSLATHAVSPNATHADGTVEYDIIHNLGYLQENATYHALLEVPNKR 616
Qy 543 PVLISRSSTFSSGKYTAHTWGDNAATNDLVYSPMLDFGLGIPMVGADICQFLGNTT 602
Db 617 PFMISRSSTFPFAGKWTGWGSDNTADWAYFESIPOAFSMGIAGLPFFGADVCGFNNSD 676
Qy 603 BELCRRMIQLGAFYFSDHSLGTTGYELRYMESVAASARKVLGLRYTLPLPYFTLMYE 662
Db 677 SELCSRWQLSGSPFFFYNNHNLGNAIDQEPVWESVAEATFSAIRYLLPYTYTLIHE 736
Qy 663 AOLNGIPIARLPFFSPDDIKTYGISSQFLGKGMVSPVLKPGVSWVTAFFRGN---- 718
Db 737 SHTTGLPIRLAFSQWQFNDRSLSGVDNQFFVGDGLVTVPLEPGVDKVKGVFPAGKEEV 796
Qy 719 WFDLPDYTRSVTASTGRVYVTLSPDPDHINVLHOGNILLAMQOKAMTTQARKTPEHLIV 778
Db 797 YYDWTY-OREVHFQXGKNETLDAPLGHIPHIRGNGVLPTEQGYTVAESRQNPFGLIWA 855
Qy 779 MSDCCASFGELFDGVEVTMGVNRGKMTFVKFIAASAKQCTIITSVDVVSGEFAVSQKW 838
Db 856 LDNDGKAQGSLLDDGESLVVDSS-----LLVSFSVSDNTLSASP--SGDYKADQF-- 904
Qy 839 IDKVTILGL 847
Db 905 LANVTILGV 913

RESULT 14
SUIS_RABIT STANDARD; PRT; 1826 AA.
AC P07768;
DT 01-AUG-1988 (Rel. 08, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sucrase-isomaltase, intestinal [Contains: Sucrase (EC 3.2.1.48);
DE Isomaltase (EC 3.2.1.10)].
GN SI.
OC Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
```

```
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86245068; PubMed=3755079;
RA Hunziker W., Spiess M., Semenza G., Lodish H.F.;
RT "The sucrase-isomaltase complex: primary structure, membrane-orientation, and evolution of a stalked, intrinsic brush border protein."
RT Cell 46:227-234 (1986).
RL
RN [2]
RP PRELIMINARY SEQUENCE OF 1-37 AND 1007-1014.
RX MEDLINE=83105704; PubMed=7152027;
RA Sjoestrom H., Noren O., Christiansen L.A., Wacker H., Spiess M.,
RA Bigler-Meier B., Rickli E.E., Semenza G.;
RT "N-Terminal sequences of pig intestinal sucrase-isomaltase and pro-sucrase-isomaltase. Implications for the biosynthesis and membrane insertion of pro-sucrase-isomaltase."
RT FEBS Lett. 148:321-325 (1982).
RL
CC -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE FINAL STAGE OF CARBOHYDRATE DIGESTION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of sucrose and maltose by an alpha-D-glucosidase-type action.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,6-alpha-D-glucosidic linkages in some oligosaccharides produced from starch and glycogen by alpha-amylase, and in isomaltose.
CC -!- SUBUNIT: THE RESULTING SUCRASE AND ISOMALTASE SUBUNITS STAY ASSOCIATED WITH ONE ANOTHER IN A COMPLEX BY NON-COVALENT LINKAGES.
CC -!- PTM: THE PRECURSOR IS PROTEOLYTICALLY CLEAVED WHEN EXPOSED TO PANCREATIC PROTEASES IN THE INTESTINAL LUMEN.
CC -!- PTM: N- AND O-GLYCOSYLATED.
CC -!- PTM: SULFATED (BY SIMILARITY).
CC -!- MISCELLANEOUS: THERE IS A HIGH DEGREE OF HOMOLOGY BETWEEN THE ISOMALTASE AND SUCRASE PORTIONS (41 % OF AMINO ACID IDENTITY) INDICATING THAT THIS PROTEIN IS EVOLVED BY PARTIAL GENE DUPLICATION.
CC -!- SIMILARITY: Contains 1 P-type (trefoil) domain.
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL: M14046; AAA31459.1;
DR InterPro; IPR000322; Glyco_hydro_31.
DR InterPro; IPR000519; P_trefoil.
DR Pfam; PF01055; Glyco_hydro_31; 2.
DR Pfam; PF00088; trefoil; 2.
DR SMART; SM00018; PD; 2.
DR PROSITE; PS00025; P_TREFOIL; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 2.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 2.
KW Multifunctional enzyme; Transmembrane; Glycoprotein; Hydrolase; Glycosidase; Repeat; Signal-anchor; Sulfation.
FT INIT MET 0
FT CHAIN 1 1826 SUcrase-ISOMALTASE, INTESTINAL.
FT CHAIN 1 1006 ISOMALTASE.
FT CHAIN 1007 1826 SUCRASE.
FT DOMAIN 1 11 CYTOPLASMIC.
FT DOMAIN 12 31 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN 32 1826 LUMENAL.
FT DOMAIN 42 59 SER/THR-RICH.
FT DOMAIN 61 108 P-TYPE.
FT DOMAIN 109 1006 ISOMALTASE.
FT DOMAIN 1007 1826 SUCRASE.
FT ACT_SITE 504 504
FT ACT_SITE 1393 1393
FT DISULFID 62 93 BY SIMILARITY.
FT 76 92 BY SIMILARITY.
```

FT DISULFID 87 105 BY SIMILARITY  
 FT MOD\_RES 390 390 SULFATION (POTENTIAL).  
 FT MOD\_RES 399 399 SULFATION (POTENTIAL).  
 FT MOD\_RES 1381 1381 SULFATION (POTENTIAL).  
 FT MOD\_RES 1384 1384 SULFATION (POTENTIAL).  
 FT CARBOHYD 41 41 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 98 98 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 454 454 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 858 858 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 895 895 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 903 903 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1001 1001 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1234 1234 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1302 1302 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1324 1324 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1339 1339 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1353 1353 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1367 1367 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1402 1402 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1534 1534 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1571 1571 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1747 1747 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1762 1762 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1798 1798 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 1826 AA; 210008 MW; 6840D03955A45BE5 CRC64;

Query Match 28.4%; Score 1354.5; DB 1; Length 1826;  
 Best Local Similarity 34.9%; Pred. No. 136-86;  
 Matches 313; Conservative 167; Mismatches 309; Indels 109; Gaps 26;

QY 35 VIGYGVKYSKVGDSGTRRSALTALPOLVKNSSVYGPDIQLLSITASLESNDRLRVITDA 94  
 DB 108 VDNHGVNVEGM---TTTSTGLEALNRKSTFTLFGNDINNVLTTESQTANRLFKLTD 164

QY 95 KHRWEIPDNLHRHQPPPPPHSLSSLYRTLLSSPTTNRKILLSHPNLSLTSLINTT 154  
 DB 165 NMRKYEVP---HQPVTBFAGPAATETLYD-----QVTEN 196

QY 155 PEGFTISKSTHDLVFDATPOPTNPNTFLIDQYLHLTSS:PGTRAHIYGLGHSRPT 214  
 DB 197 PFSIKVIRKSNRILFDSSIGP-----LVYSQDYQLQISTRLPS--EYMGFGHVKRF 248

QY 215 QLAHN---QTLTMRADIPSNPDVNLVGHSPFYMDVRSSPVAGSTGHVLLNSNGMDVE 271  
 DB 249 R--HDLVKTWPIFRDQHTDDNNNLXGHTFFMCIEDT--TKSGFVFLMNSNAMEIF 304

QY 272 YTGRI--TYKVGIGIIDLFFAGSPGVQVQFTRVIGRPAMPYAFGQOCRYGHDV 330  
 DB 305 IQPTPIVYRIGIGILDYFIELGTPTEQVQYQELIGRPAMPAYMSLGLSLRWYNSL 364

QY 331 YELQSVVAGYAKAKIPLVMMVTDIDYMDAYKDFTLDPVNFPLDKMKKVNHLKNGOKYV 390  
 DB 365 DVVKEVRNRREALPFTQVSDIDYMDKXDFYDRVAY--NLGDPFVQDLRHGOKYV 422

QY 391 VILDFGISTNK-----TYETVIRGMKHDFVLRKNG--KPYLGSVMGPGVYPPDFLKP 443  
 DB 423 ILDPALISNRASGEAYESYDRGNAQVWVNESDGTTPVGEVWPGDVTYDFTSPNC 482

QY 444 TFWTDEIKRFLNLLPVDGLWIDMEISNFI--SSPIPGSTLDNPPYKINNSGWLPII- 500  
 DB 483 EWMANEICIFHOEVNYDGLWIDMEVSSFOGSKNGCNDNTLNYPY-----IPDIV 534

QY 501 ----NKTIPPTAMHYGDIPEYVNHNLFGYLEARVTRAAIKL--TEKRPVFLSRSTSGS 554  
 DB 535 DKLMYSKTLCDMSVOYWG--KQYDVHSLVGYSMATATERAVRVPFNKRSLFTRSTFAGS 593

QY 555 KRYTAHTGDNAAATNDLVYSPISMLDFGLFPGIMVAGADICGFLGNTTEELCRRWIQLGA 614  
 DB 594 GRHAHMLGDNATATWEQMEWSITGMLFGLFGMPLVAGADICGLFAETTEELCRRWIQLGA 653

QY 615 FYPPFRSHSLGTTVOE--LYRWES--VAASARKVLGLRYTLPLPYTYLWYEAQLNGPIA 671  
 DB 654 FYPPSRNHNADGFEHQDPAFFQDSSLVVKSSRHYLNIRYTLPLPYTLFYKHAFFGETVA 713

QY 672 RPLFSFPDDIKTYGISQFLKCKGVWVSPVLKPGVSVSTAYFPRGNWFDLFDYTRSVTA 731  
 DB 714 RPLVHEFYEDTNSWVEDREFLWGPALLITPLVTOGAETVSAIYIPDAVWYD----- 764

QY 732 STG-----RYVTLSPDPDHINVHIQEGNILAMOGKAMTTQAAARKTPFHLVVMDCGA 784  
 DB 765 ETGAKRPWRKQVEMSLPADKIGLHLRGYIIPQOPAVTTTASPMNPLGLIALLDDNT 824

QY 785 SGFELFDGQVVTVMGVRGKWTFFVKFAASAKQCIITSDVSGEFAVSQKWDKVTI 844  
 DB 825 AVGDFFDDG--ETKDTVONDYILYTFVSNNNLNTCTHLYSEGTTLA----POTIKI 879

QY 845 LGLRGKTGLNGYTVRTGAVTRKGDKSLKS--TPDRKGEFFVAEISGLNLLGREFKL 900  
 DB 880 LGVTE-----TVIQVTVAENNSQSMSTHSNFTYDPSNOVLL--IENLNLGRNFRV 928

RESULT 15  
 SUITS RAT  
 ID SUITS RAT STANDARD; PRT: 1840 AA.  
 AC F23739;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Sucrase-isomaltase, intestinal [Contains: Sucrase (EC 3.2.1.48);  
 DE Isomaltase (EC 3.2.1.10)].  
 GN SI.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Intestine;  
 RX MEDLINE=95121929; PubMed=7821806;  
 RA Chandrasena G., Osterholm D.E., Sunitha I., Henning S.J.;  
 RT "Cloning and sequencing of a full-length rat sucrase-isomaltase-  
 RT encoding cDNA";  
 RL Gene 150:355-360(1994).  
 RN [2]  
 RP SEQUENCE OF 86-361 FROM N.A.  
 RC STRAIN=Fischer 344; TISSUE=Intestine;  
 RX MEDLINE=91097578; PubMed=2268340;  
 RA Traber P.G.;  
 RT "Regulation of sucrase-isomaltase gene expression along the crypt-  
 RT villus axis of rat small intestine";  
 RL Biochem. Biophys. Res. Commun. 173:765-773(1990).  
 RN [3]  
 RP SEQUENCE OF 732-1372 FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Duodenum;  
 RX MEDLINE=90381315; PubMed=2400788;  
 RA Broyard J.-P., Hugot J.-P., Perret C., Porteu A.;  
 RT "Molecular cloning and characterization of a rat intestinal sucrase-  
 RT isomaltase cDNA. Regulation of sucrase-isomaltase gene expression by  
 RT sucrose feeding";  
 RL Biochim. Biophys. Acta 1087:61-67(1990).  
 RN [4]  
 RP SEQUENCE OF N-TERMINUS OF ISOMALTASE AND SUCRASE.  
 RC MEDLINE=82167542; PubMed=6802834;  
 RX Hauri H.-P., Wacker H., Rickli E.E., Bigler-Meier B., Quaroni A.,  
 RA Semenza G.;  
 RT "Biosynthesis of sucrase-isomaltase. Purification and NH2-terminal  
 RT amino acid sequence of the rat sucrase-isomaltase precursor  
 RT (pro-sucrase-isomaltase) from fetal intestinal transplants";  
 RL J. Biol. Chem. 257:4522-4528(1982).  
 CC -I- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE FINAL STAGE OF  
 CC CARBOHYDRATE DIGESTION.  
 CC -I- CATALYTIC ACTIVITY: Hydrolysis of sucrose and maltose by an  
 CC alpha-D-glucosidase-type action.  
 CC -I- CATALYTIC ACTIVITY: Hydrolysis of 1,6-alpha-D-glucosidic linkages  
 CC in some oligosaccharides produced from starch and glycogen by  
 CC alpha-amylase, and in isomaltose.

FT	CONFLICT	222	222	C -> F (IN REF. 2).
FT	CONFLICT	230	230	L -> V (IN REF. 2).
FT	CONFLICT	240	240	T -> R (IN REF. 2).
FT	CONFLICT	252	252	G -> E (IN REF. 2).
FT	CONFLICT	283	283	L -> F (IN REF. 2).
FT	CONFLICT	294	294	G -> E (IN REF. 2).
FT	CONFLICT	326	326	T -> A (IN REF. 2).
FT	CONFLICT	350	352	VHW -> FNG (IN REF. 2).
FT	CONFLICT	359	361	YWN -> HWE (IN REF. 2).
FT	CONFLICT	732	732	S -> A (IN REF. 3).
FT	CONFLICT	735	735	E -> V (IN REF. 3).
FT	CONFLICT	841	841	E -> Q (IN REF. 3).
FT	CONFLICT	915	915	A -> T (IN REF. 3).
FT	CONFLICT	923	923	A -> R (IN REF. 3).
FT	CONFLICT	929	930	AG -> GT (IN REF. 3).
FT	CONFLICT	937	938	CR -> SQ (IN REF. 3).
FT	CONFLICT	958	961	GTCT -> ETDK (IN REF. 3).
FT	CONFLICT	979	979	Y -> C (IN REF. 3).
FT	CONFLICT	985	985	N -> H (IN REF. 3).
FT	CONFLICT	996	997	LP -> SL (IN REF. 3).
FT	CONFLICT	1009	1009	P -> A (IN REF. 3).
FT	CONFLICT	1022	1032	T -> P (IN REF. 3).
FT	CONFLICT	1026	1026	G -> E (IN REF. 3).
FT	CONFLICT	1031	1031	P -> K (IN REF. 3).
FT	CONFLICT	1093	1093	R -> S (IN REF. 3).
FT	CONFLICT	1098	1098	G -> A (IN REF. 3).
FT	CONFLICT	1301	1301	A -> D (IN REF. 3).
FT	CONFLICT	1336	1336	P -> A (IN REF. 3).
FT	CONFLICT	1338	1339	VW -> WG (IN REF. 3).
SQL	SEQUENCE	1840 AM;	210218 MW;	079ADA45E9A23E2E CRC64;

Query Match  
Best Local Similarity 27.7%; Score 1323; DB 1; Length 1840;  
Matches 298; Conservative 155; Mismatches 280; Indels 122; Gaps 27;

Qy

38 YGVKXSV-KVDSGTRRSITALPOLVKNSVSVPDIIQLISITASLESNDRLRVRTIDAKH 96

Db

121 HGNAESITNENAGKATLNRIP-----SPTLGEDIKSVILTTQTOTGNRFRFKITDPNN 176

Qy

97 RWBEIDNTLHRHQPPPPPHSLSSLYRTLSSPTTNRKILLSHPNSDLTSLINTTFP 156

Db

177 KRYEVHPQVKSETGP-----AADTLVDVQVE-----NFP 208

Qy

157 GFTISRKSTHDVLFDATPDTPNPTELFIFDIQYLHLTSSLPGTRAHYGLGHSKPTFOL 216

Db

209 SIKVIRKSNNKLCDSVGF-----LLYSNOVLQISTPLPS--EYIYFGCHIHKRFR- 259

Qy

217 AHN---OTLTMAAD-I PSSNPVNLYGSHPEVMVRSSPVAGSTHGVLLLNSGMWDVEY 272

Db

260 -HOLYWKTPWIFTRDIRPDGN-NHNDYGHQTFMGI--GDTSGKSYGVFLMSNAMEVF 315

Qy

273 TGNR-ITYYVIGIIDLYFFAGSPQQVVQEQTFRVIGRPAPMPYAFGFCQCXYGVHDVY 331

Db

316 QPTPIITYVTGGILDYIFLGDTPEQVVQOYQEVHWRPAMPAYMWNLGFQLSRWNYGSLD 375

Qy

332 ELQSVVAGYAKAPILEVNMWTDDYMDAYKDFTLDPNWFELDKMKKFVNLLHKNGOKYVV 391

Db

376 TVSEVYRRREAGIPDAQVTDIDYMEDHEKEFTYDRVKF--NGLPEFAQDLHHNG-KYII 432

Qy

392 ILDPGISYTNK----TYETVIRGMKHDFVLKXNG--KPYLGSWPWGPVYFPDFLKPSALT 444

Db

433 ILDPAISINKBANGAEYQTVRGNEKNVWVESDGTTPLIGEWPGLTYVPDFTNPQIE 492

Qy

445 FTWDEIKRFNLNLPVDGLMI DNNETSNFNISPPPIPGSTLDNPPYKINNCGVMLPIINKTI 504

Db

493 WMANECNLHQVVEYDGLWIDMNVESSFIOG-----SINLKGVLIVLN--Y 537

Qy

505 PP-----TAMHYGDIPEYNVHNLFGYLEARVTPAALIKL-TERKPFV 545

Db

538 PPTTGILDKMYSKTLCMDVAQHMG--KOYDVHSLYGYSMAIATEQAVERVFPNKRSFI 595

Qy

546 LSRSTFGSGKYTAHTWGDNAAATNDLVTSIESMLDPGLGIPMVGCADI CGFLGNTTEEL 605

Db 596 LTRSTFGSGRHHANHWLGONTASWEQMEWSITGMLFGIFGMPLVGATSCGFLADTTSEL 655  
Qy 606 CRRNIQLGAFYPPSRDSSLGTTQOE-LYRWESVAASARKVLGLRYTLPLPYFTLMYEAO 664  
Db 656 CRRMWQLGAFYPPSRHNAEGYMEQDPAYFGOD---SSRHYLTIRYTLPLPELYTLFYRAH 712  
Qy 665 LNGIFARPLPFSFPDDIKTYGISSQFLLGKGVVSPVLKPGVSVVTAAYFPRGNWEDLFD 724  
Db 713 MGETVARPFLYEFYDDINSWIEDTQFLWGPALLITPVRPGVENVSAYIPNATWIDY-- 770  
Qy 725 YTRSVTASTG-----RYVTLSAPPDHINVHIQEGNILAMQKGAMITQAAARKTFFHLLV 777  
Db 771 -----ETGIKRPWRKERINMYLPQDKIGLHLRGY-IPTQEPDVTTTASRKNPLGLIV 823  
Qy 778 VMSDCGASFGELFLDDGVEVTMGVNRGKWTFKVFAASAKQTC:ITSDVWSGEFAVSOKW 837  
Db 824 ALDDNQAAKGELFWDDG-ESKDSIEKKMYILYTFSVSNNE----LVLNCTHSSYAEGTSJ 878  
Qy 838 VICKVTILGLRKGTK 852  
Db 879 AFKTIKVLGLREDVR 893

Search completed: October 27, 2003, 10:20:32  
Job time : 17.5556 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 : Search time 23.6046 Seconds  
(without alignments)  
3678.964 Million cell updates/sec

Title: US-10-043-418-3  
Perfect score: 4768  
Sequence: 1 MKKKIPSLALGILLVFLQY.....VAEISGLNLLGREFKVLVH 903

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues  
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_76:\*  
1: Pir1:\*  
2: Pir2:\*  
3: Pir3:\*  
4: Pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4768	100.0	903	2 T09143	alpha-glucosidase
2	3243.5	68.0	913	2 J05463	alpha-glucosidase
3	2830.5	59.4	902	2 T48531	alpha-glucosidase
4	2422.5	50.8	877	2 S65057	alpha-glucosidase
5	2129	44.7	915	2 H96709	hypothetical prote
6	1956	41.0	855	2 T47534	hypothetical prote
7	1500.5	31.5	952	1 A32609	alpha-glucosidase
8	1499	31.4	864	2 JC4624	alpha-glucosidase
9	1444.5	30.3	995	2 T50267	probable family 31
10	1410.5	29.6	985	2 JC4217	alpha-glucosidase
11	1379	28.9	993	2 T38598	probable family 31
12	1374	28.8	958	1 JN0102	glucan 1,4-alpha-g
13	1348.5	28.3	1827	1 A23945	sucrose alpha-gluc
14	1323	27.7	1841	2 T10799	sucrose alpha-gluc
15	1318.5	27.7	1827	1 U0HU	sucrose alpha-gluc
16	1186	24.9	1743	2 T15893	hypothetical prote
17	1175.5	24.7	1070	2 S19686	alpha-glucosidase
18	1173	24.6	719	2 JC1200	alpha-glucosidase
19	1101	23.1	919	2 T16693	hypothetical prote
20	1041	21.8	856	2 T22575	hypothetical prote
21	825	17.3	763	2 AG1460	alpha-glucosidase
22	820	17.2	779	2 AE2402	alpha-glucosidase
23	811.5	17.0	919	2 T07391	probable alpha-glu
24	800.5	16.8	763	2 AH1097	alpha-glucosidase
25	790	16.6	818	2 AC2472	alpha-glucosidase
26	787	16.5	954	2 S46105	glucan 1,4-alpha-g
27	780	16.4	910	2 T22050	hypothetical prote
28	780	16.4	924	2 T32044	hypothetical prote
29	754.5	15.8	941	2 T32449	hypothetical prote

ALIGNMENTS

RESULT 1

T09143  
alpha-glucosidase [EC 3.2.1.20] - spinach  
C:Species: Spinacia oleracea (spinach)  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: T09143  
R:Sugimoto, M.; Furui, S.; Suzuki, Y.  
Plant Mol. Biol. 33, 765-768, 1997  
A:Title: Molecular cloning and characterization of a cDNA encoding alpha-glucosidase  
A:Reference number: Z16585; MUID:97238484; PMID:9132069  
A:Accession: T09143  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-903 <SUG>  
A:Cross-references: EMBL:D86624; NID:G2081626; PIDN:BAAL9924.1; PID:G2081627  
A:Experimental source: strain Dash  
C:Superfamily: lysosomal alpha-glucosidase; sucrose/isomaltase homology; trefoil homo  
C:Keywords: glycosidase; hydrolase  
F:145-799/Domain: sucrose/isomaltase homology <SIM>

Query Match		100.0%;	Score 4768;	DB 2;	Length 903;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 903;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MKKKIPSLALGILLVFLQYLVAGISTSENDEGVIGYKVKVKSVDGSTRSLTALPQ	60		
Db	1	MKKKIPSLALGILLVFLQYLVAGISTSENDEGVIGYKVKVKSVDGSTRSLTALPQ	60		
Qy	61	LVKNSVYGPDIQLLSITASLESNDRLVRITDAKRRWEIPDNILHRHQPPPPPHSL	120		
Db	61	LVKNSVYGPDIQLLSITASLESNDRLVRITDAKRRWEIPDNILHRHQPPPPPHSL	120		
Qy	121	SLYRTLLSPPTNRKILLSHPNSDLTFSLINTTFFGFTISRKSTHDVLFDTAPDTPN	180		
Db	121	SLYRTLLSPPTNRKILLSHPNSDLTFSLINTTFFGFTISRKSTHDVLFDTAPDTPN	180		
Qy	181	TLFIIDQYVHLTSSLPGTRAHIYGLGHSKPTFLAHNQTLTMRADIPSSNPVNLG	240		
Db	181	TLFIIDQYVHLTSSLPGTRAHIYGLGHSKPTFLAHNQTLTMRADIPSSNPVNLG	240		
Qy	241	SHPPYMDVRSSPVAGSTHGVLLNSNGMDVEYTGNRITYKVGIGIIDLYFFAGSPGQV	300		
Db	241	SHPPYMDVRSSPVAGSTHGVLLNSNGMDVEYTGNRITYKVGIGIIDLYFFAGSPGQV	300		
Qy	301	EQFTRVIGPAPMPYVAFGQCRYGHDVYBLQSVAGYAKAKIPLVMTDIDYMDAY	360		
Db	301	EQFTRVIGPAPMPYVAFGQCRYGHDVYBLQSVAGYAKAKIPLVMTDIDYMDAY	360		
Qy	361	KDFTLDVNFPLDKKKFVNHLHKQKQYVILDPGISITNKTYETIYRGMKHDVFLKRG	420		
Db	361	KDFTLDVNFPLDKKKFVNHLHKQKQYVILDPGISITNKTYETIYRGMKHDVFLKRG	420		

alpha-glucosidase  
sucrose alpha-gluc  
alpha-xylosidase (  
hypothetical 88.1  
glucosidase BH0704  
hypothetical prote  
hypothetical prote  
glycosyl hydrolase  
probable glycosyl  
alpha-xylosidase -  
probable glucosida  
alpha-glucosidase  
hypothetical prote  
hypothetical prote  
glycosidase homo  
probable glycosyl

```

QY 421 KPYLGWVPGVYVDFLKPALTEWTDKIFLNLPLVDGLWIDMNEISNFISSPIPG 480
DB 421 KPYLGWVPGVYVDFLKPALTEWTDKIFLNLPLVDGLWIDMNEISNFISSPIPG 480
QY 481 STLDNPPYKINNSGVMLPIINKTIPPTAMHYGDIPEYVNVHNLFGYLEARVTRAALIKLITE 540
DB 481 STLDNPPYKINNSGVMLPIINKTIPPTAMHYGDIPEYVNVHNLFGYLEARVTRAALIKLITE 540
QY 541 KRPVLSRSTSGSGKYTAHWTGONAAWMDLVYSIPSMDFGLFGIPMVGADICGFLGN 600
DB 541 KRPVLSRSTSGSGKYTAHWTGONAAWMDLVYSIPSMDFGLFGIPMVGADICGFLGN 600
QY 601 TTEELCRRWIOLGAFYPPSRDHSLLGTYQELRYWESVAASARKVGLGRYTLFPFYFTLM 660
DB 601 TTEELCRRWIOLGAFYPPSRDHSLLGTYQELRYWESVAASARKVGLGRYTLFPFYFTLM 660
QY 661 YEAQINGIPIARPLFPSPDDIKTYGISOQFLGKGVMSVPLKGVVSVTAYFPRGNWF 720
DB 661 YEAQINGIPIARPLFPSPDDIKTYGISOQFLGKGVMSVPLKGVVSVTAYFPRGNWF 720
QY 721 DLFDYTRSVTASTGRYVTLSPAPPDHINVIHQEGNILAMQGNAMTTOARKTPFHLLVMS 780
DB 721 DLFDYTRSVTASTGRYVTLSPAPPDHINVIHQEGNILAMQGNAMTTOARKTPFHLLVMS 780
QY 781 DCGASFGELFDGGEVTVMGVNRGKWTVPKFIASAKQTCIITSVWSGEFAVSQKWVID 840
DB 781 DCGASFGELFDGGEVTVMGVNRGKWTVPKFIASAKQTCIITSVWSGEFAVSQKWVID 840
QY 841 KVTILGLRKGTKINGYTVRTGAVTRKGDKSLKSTPDRKGFEFVAEISGLNLLGREFKL 900
DB 841 KVTILGLRKGTKINGYTVRTGAVTRKGDKSLKSTPDRKGFEFVAEISGLNLLGREFKL 900
QY 901 VLH 903
DB 901 VLH 903

RESULT 2
JC5463
alpha-glucosidase (EC 3.2.1.20) - sugar beet
N;Alternate names: alpha-D-glucoside glucohydrolase
C;Species: Beta vulgaris var. altissima (sugar beet)
C;Date: 17-Jun-1997 #sequence_revision 12-Sep-1997 #text_change 20-Jun-2000
C;Accession: JC5463; PC4330
R;Matsui, H.; Iwanami, S.; Ito, H.; Mori, H.; Honma, M.; Chiba, S.
Biosci. Biotechnol. Biochem. 61, 875-880, 1997
A;Title: Cloning and sequencing of a cDNA encoding alpha-glucosidase from sugar beet.
A;Reference number: JC5463; MUID:97321863; PMID:9178565
A;Accession: JC5463
A;Molecule type: mRNA
A;Residues: 1-913 <MAT1>
A;Cross-references: DBJ:D89615; NID:G2217947; PIDN:BA20343.1; PID:G2190276
A;Experimental source: seed; cv. NK-152
A;Accession: PC4330
A;Molecule type: protein
A;Residues: 234-261;310-365;507-541;810-840 <MAT2>
A;Experimental source: seed
A;Comment: This enzyme is an exo-glucohydrolase that catalyzes the hydrolysis of alpha-gl
C;Superfamily: lysosomal alpha-glucosidase; sucrose/isomaltase homology; trefoil homolo
C;Keywords: glycosidase; hydrolase
F:149-803/Domain: sucrose/isomaltase homology <SIM>

Query Match      68.0%  Score 3243.5; DB 2; Length 913;
Best Local Similarity 67.3%  Pred. No. 7,3e-223;
Matches 610; Conservative 122; Mismatches 166; Indels 9; Gaps 5;

QY 2 KKKIPSL---ALGILLVFLQYLVAGISTSENDPEG-VIGYGYKVKSVKVDSTGRSLTA 57
DB 3 RSKLPYICPTLAVLPLVLCWVEGATTSKNDNGEALIGYGYQVKNAKVDNSTGKSLTA 62
QY 58 LPQLVNSSVGPDPQLLSITASLESNDRLAVRTDAKRRWEIPDNLLHRHQPPPPPPH 117
DB 63 LLOLRNRPVYGPDIHFLSFTASFEEDTLIRFTDANNRRWEIPNEVLPR-PPPPSP 121

```

```

QY 118 SLSSLYRTLSSPTNRRKILLSHPNSDLTFSLINTTPPGPTISRKSTHDVLDPDPT 177
DB 122 PLSSLOHLPKPIQONQPTTIVLSHPDLSLAFTLHPTTFEFTIYRKSTHDVLDPDPT 181
QY 178 NPNTPLIFIDQYLHLTSLPGTRAHYIGLGEHSKPTFQLAHNOTLTMRAADIPSSNPDVN 237
DB 182 NPPTFLIYKQYLOLSSSLPAQAHLYLGEHTKPTFQLAHNQLLTLNADIASFNRLN 241
QY 238 LYGSHPFYMDVRSSPVAGSTHGVLLNSNGMDVEYTGNRITYKYIGIIDLYFFRAGSPG 297
DB 242 LYGSHPFYMDVRSSPMVSGTHGVFLNSNGMDVEYTGDRITYKYVIGIIDLIFAGRTPE 301
QY 298 QVVQOFTRVIGPAPMPYWAFGQOCRYGHDVYELQSVVAGYAKAKIPLEVYMTDIDQM 357
DB 302 MYLDQYTKLIGRAPMPYWAFGFHQCRWGRVDNEIETVVDKYAEARIPLEVYMTDIDQM 361
QY 358 DAYKDFTLDPVNFPLDKMKKFFVNNLHNGKQYVVDLDPGISTNKTYETVIRGMKHDFLK 417
DB 362 DAFKDFTLDPVHFPLDKMQQFVTKLHNRQYVVDLDPGINTKSYGTFRGMQSNVFIK 421
QY 418 RKGKPYLGSVWFGPVYFPDPLKPSALTFTWDEIKRFLNLLPVDGLWIDMNEISNFISSPP 477
DB 422 RGNRPYLGWVPGVYVDFLPPAARSFWVDSIKRPRDILPIDGIWIDMNEASNFITSAP 481
QY 478 IPGSTLDNPPYKINNSGVMLPIINKTIPPTAMHYGDIPEYVNVHNLFGYLEARVTRAALIK 537
DB 482 TPGSTLDNPPYKINNSGVRPINSKTIPTAMHYGNVTEYNAHNLGYFLESQATREALVR 541
QY 538 LTKRPFVLSRSTSGSGKYTAHWTGDNAAWMDLVYSIPSMDFGLFGIPMVGADICGF 597
DB 542 PATRGPFLLSRSTSGSGKYTAHWTGDNAAWMDLQYSIPTMLNFGFLGMPMIGADICGF 601
QY 598 LGNTTEELCRRWIOLGAFYPPSRDHSLLGTYQELRYWESVAASARKVGLGRYTLFPY 657
DB 602 AESTTEELCRRWIOLGAFYPPSRDHSLLGTYQELRYWESVAASARTVLGLRYELPYY 661
QY 658 TLMYEAQLNGIPIARPLFPSPDDIKTYGISOQFLGKGVMSVPLKGVVSVTAYFPRG 717
DB 662 TLMYDANLGSPIARPLSPFPDDVATYGISQFLGRGIMVSPVLOPSSIVNAYSFRG 721
QY 718 NMFDLFDYTRSVTASTGRYVTLSPAPPDHINVIHQEGNILAMQGNAMTTOARKTPFHLLV 777
DB 722 NMVLSNYSSTSSVSAGTYVLSLAPPDHINVIHQEGNIVAMQGEAMTTOARSTPFHLV 781
QY 778 VMSDCASFGELFDGGEVTVMGVNRGKWTVPKFIASAKQTCIITSVWSGEFAVSQKW 837
DB 782 VMSDHSVASTGELFDNGIEMDGGPGKWTLVRRFFAESGINNLTISSEVYVNRGYAKSQRM 841
QY 838 VIDKVTILGLRKGTKINGYTVR--TCGAVTRKGDKSLKSTPDRKGFEFVAEISGLNLLG 895
DB 842 VMDKITILGLRKRVIKEYTVQKDAGAIKVKGLGRSTSS--HNQGGFFVSVISDLRLQV 899
QY 896 REFKLVL 902
DB 900 QAPKLEL 906

RESULT 3
T48531
alpha-glucosidase 1 - Arabidopsis thaliana
N;Alternate names: protein T22P22.110
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 21-Jul-2000
C;Accession: T48531
R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancro
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24490
A;Accession: T48531
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-902 <BEV>
A;Cross-references: EMBL:AL163814

```





```

QY 596 GFLGNTTEELCRRIWLGAFYPSRDRSHSLGTYTQELYRWESVAASARKVLGLRYTL:PY 655
DB 566 GNGNTTEELCRRIWLGAFYPSRDRSHSLGTYTQELYRWESVAASARKVLGLRYTL:PY 625
QY 656 FYTLMYEAQINGIPIARPLFFSPDDIKTYGISSQFLGKGMVSPVLKPGVSVYAYFP 715
DB 626 FYTLMYEAHMTGAPIARPLFFSPDHVATYGVDRQFLGLGRVLSPVLPEGPTTVDAVFP 685
QY 716 RGNWFDLFDYTRSVTASTGRYVTLSPAPPHVNHVQEGNILAMQKAMTTOARPTPHL 775
DB 686 AGRWLYELDYSLAVATRTGKHLRVLPAATVNVHGTGTLFLQOSALTYSPARTATHL 745
QY 776 LVVMSDCGASFGELFDDGVETMGVNRKWTFFVF--IAASAKOTCIITSDWSGEFAP 833
DB 746 LVALAEDGTASGLVFLDDGDSPEYG--RRSDMSVRENYKIPNKGNAIKVSEVHNSAQ 804
QY 834 SOKWVDKVTILGLRGTGKINGYVTRGTAVTRKGDKSLKSTPDRKGEFIVAEISGLNLL 893
DB 805 SRTLVISKVVLMGHRSPAAPKLLTVHNSAEVEASSAGTRYQNAAGLGGVAHIGGLSLV 864
QY 894 LGREFKL 900
DB 865 VGEPEL 871

RESULT 5
H96709
Hypothetical protein F24J5.20 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: H96709
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizart, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maity, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Talion,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: H96709
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-915 <Sto>
A:Cross-references: GB:AE005173; NID:95734722; PIDN:AD49987.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 44.7%; Score 2129; DB 2; Length 915;
Best Local Similarity 44.8%; Pred. No. 2.2e-143;
Matches 417; Conservative 167; Mismatches 282; Indels 62; Gaps 12;

QY 7 SLALGILLVFLQYLVAGISTENDEPGVIGYGVKSVKVDSTGTRSLTALPOLVQSS 66
DB 9 AFSLSLALLILCF-----SPTQYKTIQGYRL--VSIESPDGGFGLYQVQKQK 59

QY 67 VGPDTQLLISITASLESNDRLRVRITDAKHARWEIPDNILHRHQPPPPHSLSLRYLT 126
DB 60 IYGSDDITLRLFKVHTDRLRHITDAQQRVEVPYLLPREQP-----QVGKVGKS 114

QY 127 LSSPTNRRKILLSHPNSDLTFSLLNTTPGFTISRKSTHDLVFDATPPTNPFTLFI 186
DB 115 RKSPIT-----VOEISGSELIFS-YTDDPTFAVKRSNHETLFNT-----SSLVFK 161

QY 187 DQYLHLSLPGTRAHIVGLGHSKPT-FOLAHNQTLMRAADIPSSNPVDNVLGSHPPY 245
DB 162 DQYLEISTSLP-KEASLYGLGENSQANGIKLVNPEFTLYTSDVSAINUTDLYGHPHY 220

QY 246 MDVRSFVAGSTHGVLILNSNGMDVEYTGNRITYKIVIGGIILYFFAGPSPQVVEQFTR 305

```

```

DB 221 MDLRNVCGKAYAHAVLLNSNGMDVFRGDSLYTKVIGGVDFFIAGSPPLNVDOYTO 280
QY 306 VIGRPAMPYAFGPOCRYGHDVTELQSVVAGYAKAKIPLVMWTDIDYMDAYKDFTL 365
DB 281 LIGRPAMPYWSLGFHOCRWGYNHLSVVEDVVDNYKAKIPLVDVIMDDHMDGHKDFTL 340
QY 366 DPNVFLDKMKKFFVNNLHKNQGVVILDPGISNTKTYETIRGMKHDVFLKNGKPYLG 425
DB 341 NPVAIPRAKULAFLDKIHKIMKYIIVINDPGIGVNASYGTFORAMAADVFIKYEGRFLA 400
QY 426 SVMPGPVYFDFLKPSSALTFTWDEIKRFLNLLPVDGLWIDMNEISNFS----- 474
DB 401 QVWPGPVYFDFLNPKTVMWGEIKRFDLVPIDGLWIDMNEVSNFCSGLCTIPEGKQC 460
QY 475 -SPPIG-----STLGNPPYKINNNGVMLPIINKTIPPTAMHYGDIPEYNVHN 521
DB 461 PSGEQPGWCCLDCKNITKTRWDDPPYKINATGVAPVGEFTIATSATHYNGVREYDAHS 520
QY 522 LFGYLEARVTRAAIILKTEKRPFLVSRSTFSGSKYTAHWTGDNAATWDLVYSIPSMLD 581
DB 521 IYGESETIATHKGLLNVOCKRPFILSRSTFVSGQVAAHWTGDNQGTWQSLQVSISTMN 580
QY 582 FGLFIPMGVADICGFLGNTTEELCRWIQLGAFYPSRDRSHSLGTYTQELYRWESVAAS 641
DB 581 FGIPGVPMVGSICGFPQPTTEELCNRWIEGVAFYPSRDRSHSLGTYTQELYRWESVAAS 640
QY 642 ARKVLGLRYTLFFYTYLMEYEAQINGIPIARPLFFSPDDIKTYGISSQFLGKGMVSP 701
DB 641 ARNALGMRKYKILPFLYTLNBAHMTGAPIARPLFFSPDEYTECYGNSRQFLGSSFMISP 700
QY 702 VLKPGVSVTAYPFRGNWFOLFVTRSVTASTGRYVTLSPAPPHVNHVQEGNILAMQK 761
DB 701 VLEQKTEVALFPQSGWYHMFDMTQAVSKNGKVTLPAPLNFVNVHLQNTILPTQQG 760
QY 762 AMTTOAARKTPFHLVVM-----SDCGASFGELFDDGVETMGVNRKWTFFVKFIAASAK 818
DB 761 GLISKDARTTFFSLVIAFPAGASEGVATKLYLDEDELPEMKLNGQSTVYDFVYASVNG 820
QY 819 TCIITSDVSVGEFAVSOKWIDKVTILGLRGTGKINGYVTRGTAVTRKGDKSLKST--- 875
DB 821 TMQWMSQVKGKFAISKVIEKSVLGRAGQVSEIQINGSPMTKKIEVSSKEHTYVI 880
QY 876 ---PDRKGEFIVAEISGLNLLGREFKL 900
DB 881 GLEDEENKSVMEVVRGLEMLVGKDFNM 908

RESULT 6
T47534
Hypothetical protein F16L2.150 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000
C:Accession: T47534
R:Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unsel, M.; Mewes, H.W.; Rudd, S.
submitted to the Protein Sequence Database, March 2000
A:Reference number: 224468
A:Accession: T47534
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-855 <JOR>
A:Cross-references: EMBL:AL162459
A:Experimental source: cultivar Columbia; BAC clone F16L2
C:Genetics:
A:Map position: 3
A:Introns: 69/3; 291/1; 427/3; 699/1
A>Note: F16L2.150
C:Superfamily: lysosomal alpha-glucosidase; sucrase/isomaltase homology; trefoil homol

Query Match 41.08; Score 1956; DB 2; Length 855;
Best Local Similarity 44.5%; Pred. No. 4.3e-131;
Matches 406; Conservative 146; Mismatches 259; Indels 102; Gaps 17;

```



F:140,233/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F:203-204/Cleavage site: Arg-Ala (unidentified proteinase) #status experimental  
F:390,470,492,652,925/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:518/Active site: Asp #status predicted

Query Match 31.5%; Score 1500.5; DB 1; Length 952;  
Best Local Similarity 37.9%; Pred. No. 1.6e-98;  
Matches 340; Conservative 139; Mismatches 295; Indels 123; Gaps 26;

```
QY 40 YKVKSVKSGTRRSUTLAPQLVKNSSVGP-DIQLLSITASLESNDRLVRVITDAKRR 98
Db 136 YKLENI---SSSEMGYTA--TLTRTTTFPPKDIILRLDVMETENRLLHFTIKDPANRR 190
QY 99 WEIPDNLHRHQPDPHPSLSLYRLTLSSPTNRRKILLSHPNSDLTFSLLNTTSPGF 158
Db 191 YEVLPTPHVHSRAPLSVSE-----FSEEPGV 220
QY 159 TISRSKTHVDFDTPDPTNPNTFLIFDIQYLHLSLPGTRAHYVGLGEHSPFQLAH 218
Db 221 IVRQLDGRVLLNTTVP-----LFFADQFLQSLTSLPS--QY:TGLAEHLSPMLSTS 272
QY 219 NQTLTWRAADIPSNPDVNLGSHPPYMDVRSSPVAGSTHGVLLNSNGMVEYTG-N-RI 277
Db 273 WTR-TLWNRDE--APTGCANLYGSHPEYLALED---GGSAGVFLNNSNAMDVVLQSPAL 328
QY 278 TYKVIIGGIIDLYFAGPSQGVVEQFTRVIGRPAPMPYAFGFOOCRYGYHVDVYELQSV 337
Db 329 SWRSTGGILDVIFLGEPEKSVQVQVLDVVGYPFPPYWGFLGFLHCRWGYSTATRQVV 388
QY 338 AGYAKAKIPIEVMWTDIDYMDAYKDTLPDVNPF-LDMMKKF---VNNLHKNQKYVVID 394
Db 389 ENMTRAHFPLDVQWDLDMDSRDDFT----FNKDGFRDPPAMVQELHQGRRYMMI:VD 443
QY 395 PGISTN---KYETIRGMKHDFVL-KRNGKPYLGSVWPGVYFPDFLXPSALTFTWDEI 450
Db 444 PAISSGPGAGSYRDEGLRGVFIETNETGQPLIGKWEGSTAFDPTNPTALAWMEDAV 503
QY 451 KRFNLNLPVGLWIDMNEISNFI--SSPPIGSLDNPYKINNKGWMLPIINKTIPPTA 508
Db 504 AEFDQVDPDGMWIDMNEISNFIIRGSDGCPNNELENPPV---PGVGGTLOAATICAS 560
QY 509 MHYGDIPYXVHNLFGYLEARVTRAALIKLITEKRPFVLSRSTFGSGKYTEAHWTGDNAT 568
Db 561 SHQFLTHYNLHNLGYLTAIAASHRALVKARGTRPFVLSRSTFAGHGRVAGHWTGDVSS 620
QY 569 WNDLVYSIFSMDFGLFPGIPMWGAD:CGFLGNTTEELCRWRWQLGAFYFSDHDSLGTT 628
Db 621 WEOLASSVPEILQFNLGLVPLVGADYCGFLGNTSEELCVRWTLQGLAFYFPMRHNLSLSL 680
QY 629 YQELYRWESVAASA-RKVLGLRYTLPLFYFTLMYEAQLNGCIPARLPFFSFPDDIKYGI 697
Db 681 PQEPYSFSEPAQAMRKALTLRLVALLPHLYTLFQAHVAGEIVARPLFLEPKDSS:WTW 740
QY 688 SSQFLGKGMVSPVLKPGVSVTAYFPRGNMFDLFYTRSVTA----- 731
Db 741 DHOLLWGEALLIPVLQAGKAEVTGYFPLCTWYDL--QTVPIEALGSLPPPPAAPREPAI 798
QY 732 -STGRVYTLSDPHNVHIOEQNILAMQCKAMTQAAKTPPHLLVWMSDCGASGELF 790
Db 799 HSEGWTLTAPLDTINVLRAGYIIPLOGPGLTTTESRQQPVALAVALTKGEARGELF 858
QY 791 LDGVEVTWGVNRKMTVKYFAASAKOTCIITSDV-VSGEFAVSOKWIDVKVTILG--- 846
Db 859 WDDGESLEV-LERGATQVIFL---ARNNTVNELVRTVSEAGLQ---LQKVTVLGVAT 911
QY 847 -----LRKGGKINGYVTRTGAVTRKGDGKSLKSTPDRKGEFIVAEISGLNLLGREF 898
Db 912 APOQVLSNGVPVSNFTY-----SPDTK---VLDIC-VSLLMGEQF 947
```

RESULT 8

JC4624

alpha-glucosidase (EC 3.2.1.20) - Rhizomucor circinelloides f. circinelloides

C:Species: Rhizomucor circinelloides f. circinelloides  
C:Date: 10-May-1996 #sequence\_revision 19-Jul-1996 #text\_change 20-Jun-2000  
C:Accession: JC4624; PC4149  
R:Sugimoto, M.; Suzuki, Y.  
J. Biochem. 119, 500-505, 1996  
A:Title: Molecular cloning, sequencing, and expression of a cDNA encoding alpha-glucosidase  
A:Reference number: JC4624; MUID:96271012; PMID:8830045  
A:Accession: JC4624  
A:Molecule type: mRNA  
A:Residues: 1-864 <SUG1>  
A:Cross-references: DDBJ:D67034; NID:G1498134; PIDN:BAAL1053.1; PID:G1498135  
A:Accession: PC4149  
A:Molecule type: protein  
A:Residues: 203-214:492-495;612-623;715-731;742-751;769-778 <SUG2>  
A:Note: The source is designated as Mucor javanicus IFO4570  
C:Comment: This enzyme is an exo-carbohydrase, and catalyzes the splitting of an alpha well as soluble starch.  
C:Superfamily: lysosomal alpha-glucosidase; sucrase/isomaltase homology; trefoil homol  
C:Keywords: glycoprotein; glycosidase; hydrolase  
F:122-805/Domain: sucrase/isomaltase homology <SIM>  
F:187,384,406,466,500,568,734/Binding site: carbohydrate (Asn) (covalent) #status pred  
F:430/Active site: Asp #status predicted

Query Match 31.4%; Score 1499; DB 2; Length 864;  
Best Local Similarity 38.2%; Pred. No. 1.7e-98;  
Matches 324; Conservative 147; Mismatches 280; Indels 98; Gaps 18;

```
QY 12 ILLVFLQYLVLVAGISTSENDEPGVIGY---GYKVKV---VKVDSGTRRLTALPOLVKNSS 66
Db 4 VSFIFVAIALITGNVLQCOTDATYAVSSSAPGYKIDGHVRKTEAGLHPLTLNRSNGKKTG 63
QY 67 V--VGPDIOQLLSITASLESNDRLVRITDAKRRWEIPDNLHRHQPDPHPSLSLYR 124
Db 64 IDTGKTKIDITVDVEYETERLHVKISDKAKQYLVDPSPGLGERP----- 110
QY 125 TLLSSPTNRRKILLSHPNSDLTFSLLNTTFFGFTLSRSKSTHVDLPDTPDPTNPTFLI 184
Db 111 -----QIKHVVSFKHSLNDFQ-YTAKPFSPKVKRDKDKTIFDTNMP-----LV 154
QY 185 FIDQYLHLTSSLPTRAHYVGLGEHSPFQLAHQTLTWRAADIPSNPD---VNLVYS 241
Db 155 PEDQYELSLTKVP-EDANIYIGIEVTAP-FRTRHNTV-TLWARD---NPDDEFYRNYGA 207
QY 242 HPFYMDYRSPVAGSTHGVLLNSNGMDVEYTGRTYKVIIGGIIDLYFAGPS--PGQV 299
Db 208 HPYIQEVRD---GKAHGALLMNAHGMVITTEGRITIKVIGIGILDFFPAPKSGKNDL 263
QY 300 VEQFTRVIGRPAPMPYAFGFCQCRYGHDVYELQSVVAGYAKAKIPIEVMWTDIDYMDA 359
Db 264 SIAYTDLIGKPMPSHMLGWHCHRCYGPNIKDVETVKRYKEANIPLQTVWVIDDYME 323
QY 360 YKDTLPDVNPFPLDKMKKFFVNNLHKNQKYVVIDLPGISNTKTYETVIRGMQDVPFKR- 418
Db 324 TKDFTFKVNFQDPMITGLGEQLHKQOQNYVMVDPAISANTTYEYVVRGTEMDVWIKNA 383
QY 419 NGKPYLGSVWPGVYFPDFLKPSSALTFTWTDIEIKRFLNLLPVDGLWIDMNEISNF----- 472
Db 384 DGSDFIGSWPGFTTFPDWHHPNATKYWKIEIDFVDMGLVDGLWIDMNEPASCCLSGC 443
QY 473 -----ISSPPIPGSTLDNPYKINNKGWMLPI 499
Db 444 SGKVDAGNOPYRWYTYTEEEQAANHTRWEKELKAMGNPGEERNLLYPKAIINNAGNLS- 502
QY 500 INKTIPPTAMHYGDIPEYNVHNLPGYLEARVTRAALIKLITEK-RPVLRSSTFGSGSKYT 558
Db 503 -EPTVATLALYGNIPHYDIHNLGYHABSHITQALIKHKNKIRPFVLTRSSPFGSKSV 561
QY 559 AHWTDGDNAAWTNDLVYSIPSMDFGLFPGIPMWGADICGFLGNTTEELCRWRWQLGAFYF 618
Db 562 GHWTDGDNHSPYPLKNSIANILNFMQFVGVSADVCGFNSDITTEELCTRWELICAYFP 621
QY 619 SRDHSLSGTTTYQELYRWESVAASARKVLGRYTLPLPYFTIMYEAQLNGIPIARPLPFSF 678
Db 619 SRDHSLSGTTTYQELYRWESVAASARKVLGRYTLPLPYFTIMYEAQLNGIPIARPLPFSF 678
```

Db 622 ARNRNRNAKDQEPYLNVESTAEASRIAINTREYEMLPYFYLFEESNLGLGWRPLIFEY 601  
QY 679 PDIKTYGISSQFLGKGMVSPVLKPGVSVTAFFPRGNMFLDFYTRSVASTG--R 735  
Db 682 PAYEELYSNDVTLVSGDILLSPVLDEKTSVKAQFGQWYDWTHELTVDNKSNKKVK 741  
QY 736 YVTLSAPPDHNVHIOEENILAMOGKAMTTOAARKTPFHLLVWMSDCGASFGELFLDDGV 795  
Db 742 TVTLDAPLTHPIHRCGAIIPTKTPKYVTGETATPYNLVIADLKKGQASGRLYIDDE 801  
QY 796 EVTMGVNRG 804  
Db 802 SLEVKSSG 810

RESULT 9  
T50267  
probable family 31 glucosidase [imported] - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 21-Jul-2000  
C:Accession: T50267  
R:Hunt, C.; Aves, S.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, December 1999  
A:Reference number: Z25031  
A:Accession: T50267  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-995 <HUN>  
A:Cross-references: EMBL:AL133522; PIDN:CAB63549.1; GSPDB:GN00066; SPDB:SPAC922.02c  
A:Experimental source: strain 972h(-); cosmid c922  
C:Genetics:  
A:Gene: SPAC1039.11c; SPDB:SPAC922.02c  
A:Map position: 1  
C:Superfamily: Schwannomyces glucan 1,4-alpha-glucosidase GAM1; sucrase/isomaltase homolog

Query Match 30.3%; Score 1444.5; DB 2; Length 995;  
Best Local Similarity 33.8%; Pred. No. 1.6e-94;  
Matches 339; Conservative 151; Mismatches 291; Indels 223; Gaps 29;

QY 1 MKKIPSLALGILLVFLQV--VAGISTSEND-----P 32  
Db 6 MKNLNLASLFLSFFLLACEFTFSVADFTSTSDATHPSATATEDIFSTPAVPSGLAQN 65  
QY 33 EGVIGY-----GYKVKSVKVDSTRSLTALPOLVKNSS-VYGPDIQLLSITASLESND 85  
Db 66 SVYEPYRGDKCGYNAIQV---SEYKGVLAQLQNGDPCVAYGTDYPPFLALNVSPSID 122  
QY 86 RLVRITDAK-----HRRWEIPDNLHRRHQPFPFPHSLSSLYRTLSSFTNRRKILL 139  
Db 123 RLHVSIOQLYGAQFQFQSKRTDWDAPLYHFQ-----154  
QY 140 SHPNSDLTFSL-INTTPFGFTISRKSTHDLVLFDAATPDTPNTPNTEFLIFIDVYHLTSSLPG 198  
Db 155 --QFGRDRTYFNSQPFQFWTVRSQGEVLFDTRGHK-----LIFEDQVIELTNNVD 206  
QY 199 TRAHYGLGE--HKKPTFQLAHOTLTLRAADIPSSNP---DNLVXGHPFYMDVRSSP- 252  
Db 207 D-VYNYGLAETVHG---LRLGNLTRLTFA---NGNPTPLDRNAYGTHPFLYHRYTPS 258  
QY 253 -----VAGSTHGVILLNSNGMDEVYGNRITVYKVGIIIDLIFPAG--PSPGVV 300  
Db 259 ENLNSDQSPVTSSTHGVMLTANGMEVLRPNYLOYRIIGVVDLYIYVGGTKNPKNDV 318  
QY 301 EQFTRVIGRAPMPYPAFGFQOCRYGHDVYVELQSVVAGYAKAKIPLVEMVMTDIDMAY 360  
Db 319 SQFVQSVGTPAHQWHTFGFHICRWGVNVPDLVEUVENKFNFEIPVDTPWSDIDIMYEY 378  
QY 361 KOFTLDPVNPFPLDKKKFVNNLHKNQKQYVYILDPGI-----STNKTYYTIRGMKD 413  
Db 379 RDTVESNAFPKDKMMEFFNSLQSNQHYPYIIDAIIYAANPINRSDVYPPYVEGVRRD 438  
QY 414 VFLKRNKGP-YLGSVWVGVPYFPFLKPSALTFTWDEIKRPLN-----LPVQG 461

Db 439 IFLRNPDRSLRYGVNWPFGFTTPTPTFTNPTTNYWTECL-----MNLAAFGYNSFFPLPSG 495  
QY 462 LMDWNEISNF-----472  
Db 496 LMDWNEPTSFICSGCGTDKLDQNPVHPAFILEGEPNNVMYMPGFEHTNASEHASAYO 555  
QY 473 -----ISSPPI---POSTLDNPPYKINNNGVMLPIINKTIPPTAMHY 511  
Db 556 ASVSQYATATSTVESVKATSTPLNVRPKYNNYPPYALNTEQGEGLSNLGVSNATYH 615  
QY 512 GDPEYNVHNLFGYLEARVTRAALIKLEK-RPFVLSRSTFGSGKYTAHWTGDNAATWN 570  
Db 616 DGTVRNLENTYGDQSRVYDLSIENVRPFIILSRSTFVSGSKYAAHMLGDNVSLWS 675  
QY 571 DLVYSIPSLDFGLGIPMGVADI CGFLGNTTEELCRRWIQLGAFYPSRDRHSSLGTYO 630  
Db 676 NMIFSIPGALTNNMGLPMVGADVCGFMGNTDELCRMMALGAFPLFYRNHNSIGSIQ 735  
QY 631 ELYRWESVAASARKVLGURTYLLPYFTYLMYEAQLNGIPIARPLFFSPDDIKTYGISSQ 690  
Db 736 EPYRWESVAESSRCAMNIRYSLPYWYTLMYEASSOGLPIRLPFEFEPNEPSLANAD 795  
QY 691 FLGKGWVSPVLKPGVSVTAFFPRGN---WFDLFYTRSVTASTGRVYTVLSAPDHIN 747  
Db 796 FMVGSALLVTPLEPNVDYVRGVFPDKNSTIWDYWDH-KVIYRQHNENITLSAPLTHIN 854  
QY 748 VHIQENILAMOGKAMTTOAARKTPFHLLVWMSDCGASFGELFLDDGVETMGVNRGKWT 807  
Db 855 VAIRGNIIIPMKQSLTTHETKQNPIDLLVALDSRKACGSLYDDGV---SIQESTL 910  
QY 808 FVKFIAASAKOTCIITSDVVS---GEFAVSQKMWIDKVTILGL 847  
Db 911 FVKFVA-----NGDSLSIESYGDILQVHEP--LSKTIILGL 943

RESULT 10  
JC4217  
alpha-glucosidase (EC 3.2.1.20) - Aspergillus oryzae  
C:Species: Aspergillus oryzae  
C:Date: 21-Nov-1995 #sequence\_revision 08-Feb-1996 #text\_change 21-Jul-2000  
C:Accession: JC4217  
R:Minetoki, T.; Gomi, K.; Kitamoto, K.; Kumagai, C.; Tamura, G.  
Biosci. Biotechnol. Biochem. 59, 1516-1521, 1995  
A:Title: Nucleotide sequence and expression of alpha-glucosidase-encoding gene (agda) ;  
A:Reference number: JC4217; MUID:96032211; PMID:7549103  
A:Accession: JC4217  
A:Molecule type: DNA  
A:Residues: 1-985 <MIN>  
A:Cross-references: DDBJ:D45179; NID:gi054564; PIDN:BAA08125.1; PID:gi054565  
C:Comment: This enzyme catalyzes the liberation of alpha-glucose from the non-reducing  
portant enzyme in the food industry as the isomalto-oligosaccharides.  
C:Genetics:  
A:Gene: agda  
A:Introns: 234/2; 371/2; 428/2  
C:Superfamily: Schwannomyces glucan 1,4-alpha-glucosidase GAM1; sucrase/isomaltase hor  
C:Keywords: glycoprotein; glycosidase; hydrolase  
F:152-898/Domain: sucrase/isomaltase homology <STM>  
F:126-145,255,349,424,508,536,539,602,624,661,835,881,929,957/Binding site: carbohydrate  
F:492/Active site: Asp #status predicted

Query Match 29.6%; Score 1410.5; DB 2; Length 985;  
Best Local Similarity 34.3%; Pred. No. 4.3e-92;  
Matches 333; Conservative 153; Mismatches 291; Indels 195; Gaps 29;

QY 21 LVAGISTSENDPEGV-----IGYKVKSVKVDSTRSLTALPOLVKN-SSVYGGPIQLL 75  
Db 49 LVANI-----DDFQAVNAQSVCPGYKASDVYKSS---QGFASLELAGPCNVYGTVD 101  
QY 76 SITASLESNRLVRIT---DAKRRWEIPDNLHRRHQPFPFPHSLSSLYRTLSSPT 131  
Db 102 TLTVQYQAKRLNIQIVPTYFDASNASWYI-----LSEELVPRPK 141  
QY 132 TNRRKILLSHPNSDLTSLINTTPTFGFTISRKSTHDLVLFDAATPDTPNTPNTEFLIFIDQVILH 191

```
Db 142 ASQN---ASVQSDFWVMSNPSFNFKYRKATGDVLFN-----TKGST-LVYENQIE 192
Qy 192 LTSSLPGRTRAHYIGLGEHSKPTFQLAHNQTLMRAADIPSSNP-DVNYLGHSPFFMDVRS 250
Db 193 FVTLPL-BEYNLYGLGERMN-QLRLENANLTLYAADI--ADPIDONTYGHAFYLDTRY 248
Qy 251 SPVAGS-----THGVLLNSMGMDVEYTCGNRTYKVIKGGIIDL 289
Db 249 YKVGQNKSHITVKSSEAPSOEYYSYSGVFLRNAHGEILLROKLIWRTLGSDVLT 308
Qy 290 FRAGSPGQVTEQF-TRVIGRPAPMPYAFGQOCRYGHDVYELQSVVAGYAKAKIPLE 348
Db 309 FYSGPTQAEVTRYQYQLSTVGLPAMQOYNLTGFHQCRGWYNMSEFEDVLANFEREIPLE 368
Qy 349 VMWTDIDYMDAYKDFTLDPVNFLOKMKFVNHLKNGKYYVILDPGI-----SNTKT 402
Db 369 YLWADIYMHGRNFNDQHRFSYEGEKFNLKHAGGRWVPYVDGALYIPNPENASDA 428
Qy 403 YETIRGKMDHDFLKR-NGKPYLGSVMPGPVFPDFLPKPSALTFTWDEIKRFLNLLPVDG 461
Db 429 YETDRGAKDDVFIKPYDGSLSLIGAVPGYTYVPDMHHPKASDFWANELVTWNKLYHDG 488
Qy 462 LMIWNNEISNF----- 472
Db 489 VMYDMAEVSSFCVSGCGTGNLSNPNAPHPFALPGEPGNVYDYPEGFINITNATEASASA 548
Qy 473 -----ISSPPIG-STLDNPPYKINNSGVMLPIINKTIPPTAMHYGD 513
Db 549 GAASGAAASSTTTAPYURTTPTGVRNVDRPPYVNIHVQGHDLNVHAISPNSTHSDG 608
Qy 514 IPEYVHNLFGLYLEARVTRAALIKLTE-KRPVLSRSTFGSGKYTAHWTGDNAAATNDL 572
Db 609 VQEDVHSLYHGGINATYHGLLKVENKRPILARSTFGSGKWAGHMGDGNFSKWGM 668
Qy 573 VYSISMLDFGLFGIPMGADTCGFLGNTTEELCRRWIOLGAFYFSDHDSLSGTYQEL 632
Db 669 FFSISQALQSLFGIPMGFVDTCGNGNTDEELCNRMOLSAFFPYRNHNLSAIPQEP 728
Qy 633 YWESVAASARKVGLRYLLPYEYTLMEYLAQLNGIPIARPLFFSPDDIKTYGSSOFL 692
Db 729 YRWASVIDATKAAMNRIYAILFYFLAHTTGSTVNRALAMEFPDPSLAAVGTQFL 788
Qy 693 LGKGVMVSPVLKPGVVSVTAYFP---RGN-WFDLFDYTRSVTASTGRVYTLSPDHLNV 748
Db 789 VGPSVWVIVLEPQVDTVQGVFGVGHGEWYDWYSQT-AVDKPGVNTTISAPLGHIPV 847
Qy 749 HQEGNILAMQKAMTTOARKTPPHLLVVMDCGASFGELFLDDGVEY---TMGVNRRG 804
Db 848 FVRGGSILPMQEVALTTRDARKTPMSLLASLSSNGTASGQLYLDGSEYYPEDTILSDV-- 905
Qy 805 KWTFFVKFIA-----ASAKQTCIITSDVVSGBFAVSQKW-----VIDKYTILGLRK--GT 851
Db 906 -----FLASRSTLRASGT-----WKEANPLANVTVLGVTEKPSV 942
Qy 852 KINGYVTRTCAV 863
Db 943 TLNGETLSSDSV 954

RESULT 11
T38598
probable family 31 glycosyl hydrolase (alpha glucosidase) precursor - fission yeast (sch
C:Species: Schizosaccharomyces pombe
C:Date: 20-Oct-2000 #sequence revision 20-Oct-2000 #text_change 20-Oct-2000
A:Accession: T38598; S62559; T38911
R:Pearson, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, November 1995
A:Reference number: Z21801
A:Accession: T38598
A:Molecule type: DNA
A:Residues: 1-385 <PEA>
A:Cross-references: EMBL:Z67961; PIDN:CAA91887.1; GSPDB:GN00066; SPDB:SPAC30D11.01c
```

```
A:Experimental source: strain 972h-; cosmid c30D11
R:Pearson, D.; Churcher, C.M.
submitted to the EMBL Data Library, November 1995
A:Reference number: S62559
A:Accession: S62559
A:Molecule type: DNA
A:Residues: 1-384 <PE2>
A:Cross-references: EMBL:Z67961
R:Pearson, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z21817
A:Accession: T38911
A:Molecule type: DNA
A:Residues: 352-993 <PE3>
A:Cross-references: EMBL:Z67978; PIDN:CAA93572.1; GSPDB:GN00066; SPDB:SPAC56F8.01
A:Experimental source: strain 972h-; cosmid c56F8
C:Genetics:
A:Gene: SPDB:SPAC30D11.01c; SPDB:SPAC56F8.01
A:Map position: 1L
C:Superfamily: Schwanniomyces glukan 1,4-alpha-glucosidase GAML; sucrase/isomaltase ho
F:165-384/Domain: sucrase/isomaltase homology (fragment) <SIM>

Query Match: 28.9%; Score 1379; DB 2; Length 993;
Best Local Similarity 33.3%; Pred. No. 7.7e-90;
Matches 312; Conservative 170; Mismatches 287; Indels 168; Gaps 25;

Qy 22 VAGISTSENDEPGVIGYGVKVKVSDGTRRSLTALPOLVKNSS-VYGPDIQLLSITAS 80
Db 67 VAEPTLYESSRGLSCPGYQARNI---SEYSYGLAILLADGACVAYGIDYPLYLNV 123
Qy 81 LESNDLRVRITDAKHRRWEIPDNILHRHQPPLPPHSLSSLYKTLSSPTTNRKIL-- 138
Db 124 YDTEERVHISLDLNTQFOL-----SNRQVMDA 153
Qy 139 LSHPNSDLTFSL-----INTPFGTTSRKSTHVDLFDATPDPPTNPNTFLIFDQYLHL 192
Db 154 PLFYRSNFGNLQYNFSNTDPEFWITRIADQVLDTRGNP-----LIFEDQYIEL 207
Qy 193 TSSLPGRTRAHYIGLGEHSKPTFOLAHNQTLMRAADI PSSNPVNLYGSHPPFYMDVRSSP 252
Db 208 TTNMV-EDYNYGU-SGSQSFRLGNLTKTFWATGY-SDSPEANMYGSHPPFYMEQRIY 264
Qy 253 V-----AGSTHGVLLNSMGMDVEYTCGNRTYKVIKGGIIDLFPFAGP--SPQGVVEQFT 304
Db 265 IGTNTYTSASHGVMLSSNGMEVLLRSTYIKYRMIGIIDLFFVSGSTVSPKYTIQYV 324
Qy 305 RVIGRPAPMPYAFGQOCRYGHDVYELQSVVAGYAKAKIPLEVMWTDIDYMDAYKDF 364
Db 325 QSIGTPTMQPYMSLGLGFQMSRWGYKTLSDLINMRSYLNASNIPTGEFMDIDYMSERTFT 384
Qy 365 LDPVNRPLDKMKKFVNHLKNGKYYVILDPGI-----STNKTVETYIRGMKHDFVLK 417
Db 385 VNSTAFPNTQLDFFRSLDSEHQHYVPVLDPAIYAANPNKASDRTTYPPYSGFEDNIFIK 444
Qy 418 R-NGKPYLGSVMPGPVFPDFLPKPSALTFTWDEIKRFLN-----LPVDGLWID 465
Db 445 NPNGSAVYGMWPGFVYVPDFTNPVAVLYQWKQGI---LNLSTAFGSNYSDLPFSGCLD 501
Qy 466 MNEISNF-ISS-----PP-IPGS----- 481
Db 502 MNEPTSFICSGSDLLKLNLFVHPFSLPGVDGNKYSPEDFNATNTTEYKSVSRASQS 561
Qy 482 -----TLDPNPPYKINNSGVMLPIINKTIPPTAMHYGDI 514
Db 562 QYKATATSEKSHETPSESSESLINKPFSINYPYALDQVLTETDHLAQFGVSPNATWGN 621
Qy 515 PEYNVHNLFOYLEARVTRAALIKLTEK-RPFLVSRSTFGSGKYTAHWTGDNAAATNDLV 573
Db 622 LRYNLFNTYGVSESKISFEALNSIQPNIRPFLLSRSTFVSGRYAAHAWLGDNKSQWDMV 681
Qy 574 YSTPSMLDPLGFI PMVGADICGFLGNTTEELCRRWIOLGAFYFSDHDSLSGTYQELY 633
Db 682 SSISLITFLLGI PMVGADVCGNGTDEELCARWALGAFLPFYRNHNSLGSIPQEPF 741
```

```

QY 634 RMESVAASARKVLGRYTLTPYFYLYMIAQLNGIPIARPLFFSPDDIKTYGISSOFL 693
  ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : |||
Db 742 RAASVAEASRSALIRYSLPYYLMTHTASVDGTPMVRPLFFEPKO:SLASVDKQFM 801
  ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : |||
QY 694 GKGVMVSPVLKPGVSVTAFFRGN---WFDLFDYTRSVTASTGRYVTLSPAPPDHVHI 750
  ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : |||
Db 802 GTALLISPALEPNTTVIOGIIIPGNDTIWYDYNHS-VINHVDENITMSAPLGVNIAV 860
  ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : |||
QY 751 QSGNILLAQGKAMTQAAKTKPFLHLLVMSDCGASFGELFDDGVETMGVNRGKWTVK 810
  ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : |||
Db 861 RGNIIIPLOQPGVTVYESRNNYSLLIANNNGFASGSLYIDDG--ISMQNSS--LSVK 916
  ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : |||
QY 811 FIAASAKOTCIITSDVWSGEFAVSQKWKIDKVTILGL 847
  ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : |||
Db 917 LNSNSNTITCVSGTWSSP-----SLANTITILGL 946
  ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : |||

RESULT 12
JN0102
Glucan 1,4-alpha-glucosidase (EC 3.2.1.3) GAM1 precursor - yeast (Schwanniomyces occiden
N:Alternate names: acid maltase; glucoamylase
C:Species: Schwanniomyces occidentalis
C:Date: 10-Mar-1994 #sequence_revision 05-Apr-1995 #text_change 18-Jun-1999
C:Accession: JN0102
R:Dohmen, R.J.; Strasser, A.W.M.; Dahlems, U.M.; Hollenberg, C.P.
Gene 95, 111-121, 1990
A:Title: Cloning of the Schwanniomyces occidentalis glucoamylase gene (GAM1) and its exp
A:Reference number: JN0102; MUID:91071592; PMID:1979298
A:Accession: JN0102
A:Molecule type: mRNA
A:Residues: 1-958 <DOH>
A:Cross-references: GB:M60207; GB:M34666; NID:g169845; PIDN:AAA33923.1; PID:g169846
C:Genetics:
A:Gene: GAM1
C:Superfamily: Schwanniomyces glucan 1,4-alpha-glucosidase GAM1; sucrase/isomaltase homo
C:Keywords: glycoprotein; glycosidase; hydrolase; polysaccharide degradation
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-958/Product: glucan 1,4-alpha-glucosidase #status predicted <MAT>
F:156-876/Domain: sucrase/isomaltase homology <STM>
F:61,78,197,403,416,513,580,602,813,907/Binding site: carbohydrate (Asn) (covalent) #sta
F:470/Active site: Asp #status predicted

Query Match 28.8%; Score 1374; DB 1; Length 958;
Best Local Similarity 33.8%; Pred. No. 1.7e-89;
Matches 328; Conservative 142; Mismatches 315; Indels 184; Gaps 27;

QY 3 KKIPSLAGILLAVFLLOYL-----VAGISSENDPEGVI----- 36
  ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : |||
Db 5 KLIKSIIVIGLVLVAIQAPASSIGSSASSSSSQATIPNDVTLGVKQIPNIFNDSA 64
  ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : |||
QY 37 -----GYGKYSKVKVDSGTRRSUTALPOLVKNSSVYGPDIQLLSITASLESNDRLRVR 90
  ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : |||
Db 65 VDANAAAKGYDLVNV---TNTRGTLGILKXEAINIYGYDFDLNLIVEYQADTRLNVH 121
  ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : |||
QY 91 ITDAKHRRWEIPDNIHRHQPPPPPHSLUSLY-----RTLLSSPTNRRKILLSHNSDLT 147
  ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : |||
Db 122 I-----EPTDLSDFVFLPEHLVWVKPLVEGDAQSYNFDSNLSLV 158
  ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : |||
QY 148 FSLINTTPFGFTISRSKTHDVLFDATPDPTNPNTFLIPIDQVLYLCTSSLPGTGRAHIYGLG 207
  ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : |||
Db 159 FEYSN-TDFSFEVIRSSKTEVLFTKGNP-----LVFSNQFIQNSSLPKNHV-ITGLG 210
  ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : |||
QY 208 EH-----SKPTQLAHNOITLRAADIPSSNP-DVNLYGSHFPYMDVRSSPVAGSTHGLV 261
  ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : |||
Db 211 ESIHGLVNEP-----GSVKTLFANDV---GDPIDGNIYGVHPVILDQRYD--TETHAVY 260
  ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : |||
QY 262 LNSNGMVEYTGNRITYKVIIGIIDLFPAGSPQGVVEQTRVIGRAPMPYWAFFGQ 321
  ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : |||
Db 261 WRTSAIQEVLIGESITWRALSGVIDLYFFSGPTPKDAIQQYVKEIGLPAFQPYNSLGVH 320
  ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : |||
QY 322 QCRYGVDHYELQSVVAGAKAPILEVWNTDIDYMDAYKDTLPVNFPLDKMKFFVN 381
  ||| ||| : : : ||| ||| : : : ||| ||| : : : ||| ||| : : : |||

```

## RESULT 13

```

A23945
sucrose alpha-glucosidase (EC 3.2.1.48) / oligo-1,6-glucosidase (EC 3.2.1.10) - rabbit
N:Alternate names: small intestinal sucrase/isomaltase (SI)
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 24-May-1996
C:Accession: A23945; B25987; A29163
R:Hunziker, W.; Spiess, M.; Semenza, G.; Lodish, H.F.
Cell 46, 227-234, 1986
A:Title: The sucrase-isomaltase complex: primary structure, membrane-orientation, and
A:Reference number: A23945; MUID:86245068; PMID:3755079
A:Accession: A23945
A:Molecule type: mRNA
A:Residues: 1-1827 <HUN>
R:Sjoestrom, H.; Noren, O.; Christiansen, L.A.; Wacker, H.; Spiess, M.; Bigler-Meier,
FEBS Lett. 148, 321-325, 1982
A:Title: N-terminal sequences of pig intestinal sucrase-isomaltase and pro-sucrase-iso
A:Reference number: A25987; MUID:83105704; PMID:7152027
A:Accession: B25987
A:Molecule type: protein
R:Frank, G.; Brunner, J.; Jauser, H.; Wacker, H.; Semenza, G.; Zuber, H.
FEBS Lett. 96, 183-188, 1978
A:Title: The hydrophobic anchor of small-intestinal sucrase-isomaltase. N-terminal seq
A:Reference number: A29163; MUID:79086207; PMID:729784
A:Accession: A29163
A:Molecule type: protein
A:Residues: 2,'VNA',6-32,'XXX',36-38,'1010-1014','E' <SJ2>
C:Comment: Carbohydrate analysis of the mature enzyme complex indicates both N- and O-
C:Complex: the two product chains remain associated after cleavage

```

C;Function: <ISM>  
A;Description: oligo-1,6-glucosidase catalyzes the hydrolysis of 1,6-alpha-D-glucosidic  
A;Pathway: carbohydrate digestion  
C;Function: <SUC>  
A;Description: sucrose alpha-glucosidase catalyzes the hydrolysis of the 1,2-alpha-D-glu  
A;Pathway: carbohydrate digestion  
C;Superfamily: sucrose/isomaltase; sucrose/isomaltase homology; trefoil homology  
C;Keywords: carbohydrate digestion; duplication; glycoprotein; glycosidase; hydrolase; m  
F;2-1007/Product: oligo-1,6-glucosidase (isomaltase chain) #status experimental <ISO>  
F;13-32/Domain: membrane associated #status predicted <TM>  
F;43-65/Region: serine/threonine-rich  
F;189-840/Domain: trefoil homology <TRF1>  
F;931-977/Domain: sucrose/isomaltase homology <SIM>  
F;1008-1827/Product: sucrose alpha-glucosidase (sucrose chain) #status experimental <SUC  
F;1062-1734/Domain: sucrose/isomaltase homology <SIM>  
F;12/Binding site: carbohydrate (Thr) (covalent) #status experimental  
F;99,455,859,896,904,1235,1303,1325,1340,1354,1368,1403,1535,1572,1748,1763,1799/Binding  
F;1007-1008/Cleavage site: Arg-116 (trypsin) #status predicted

Query Match 28.3%; Score 1348.5; DB 1; Length 1827;  
Best Local Similarity 34.7%; Pred. No. 3.1e-87;  
Matches 312; Conservative 167; Mismatches 310; Indels 109; Gaps 26;

QY 35 V-GYGVKVS-KVDSGTRRSALTALPOLKNSVYGPDIQLLSITASLESNDRLVRITDA 94  
Db 109 VDNHGYNVEGM---TTTSTGLEARLNKSTPTLFGNDINNVLTTESQANRLRFLKLTDP 165  
QY 95 XHRRWELPNILHRHQPPTPPHSLSSLYRTLLSSPTTNRKILLSHPNSDLTFSINTT 154  
Db 166 NKKRYEVP---HQVFTEFAGPAATETLDV-----QVTEN 197  
QY 155 PFGFTSRKSTHVDLPDTPNPNTFLIFIDQYLHLTSSLPGRHAIYGLGHSKPTF 214  
Db 198 PFSIKVIRKSNRILDFSSIGP-----LVSDQYLQISTRLPS--EYWGGEHVRHFR 249  
QY 215 QLANN---QTLTNRADIPSSNPVNLVYSHPPYMDVRSSPVAGSTHGVLLNSNGMDVE 271  
Db 250 R--HDLVMTWPIFTTRDEIPGDN--NENLYGHOTFFMGI--GDTSGKSYGVFLMNSNAMEVFI 305  
QY 272 YTGNRIT-YKVIIGIIDLFFAGSPGVQVQFTRVIGRPAHPMPYAFQOCRYGHDV 330  
Db 306 IQPTPIVTVRVIGIIDLFFAGSPGVQVQFTRVIGRPAHPMPYAFQOCRYGHDV 330  
QY 331 YELQSVVAGYAKAKIPLVWMTDIDYMDAYKDTLPDVPNEPLDKMKKFNVLHKNQKQYV 390  
Db 366 DVVKEVVRNRREALIPDQVSDIDTMEKDKQFYDORVAY--NGLPDPFQDLHDHGQKV 423  
QY 391 VILDPGISTNK-----TYETIRGMKHDVFLKRG--KPYLGSVWPGPVYFPDFLKPSAL 443  
Db 424 IILDPAISINRRASGEAYESYDRGNAQVWVNESDGTTPVGEVWPGDTPVDFTPNCI 483  
QY 444 TFWTDEIKRFLNLLPVDGLWIDMNEISNFI--SSPIPGSTLDNPPYKKNNGVWLPIT- 500  
Db 484 EWWANECNIFHOEVNDGLWIDMNEVSSVQSGKNCNDNTLNPY-----IPDIV 535  
QY 501 -----NKTIPPTAMHYGDIPEYVNHMLFGYLEARVTRAALIKL--TEKRPVFLSRSTFSGS 554  
Db 536 DKLWYSKTLCDMSQVYWG--KQYDVHSLYGSMAIATERAVERFVFNKRSPILTRSTFAGS 594  
QY 555 GKTYAHTWGTNAATWDLVYSIPSLMDPLFGIPIMYGADICGFLGNTTBELCRRWQLGA 614  
Db 595 GRAAARWLGNTATWEGMWSITGMLEFGLPGPLVGADICGFOATTEELCRWMLGA 654  
QY 615 FYPFSRDHSLGTYQE--LYRNES--VASARKVLGRITLLPYFTLMEYQLNGIPIA 671  
Db 655 FYPFSRNHNDAGFHEQDPAFFQDSDLVKSRRHYLNIRYTLPLFLYKAAHGETVA 714  
QY 672 RPLVFFPDDIKTYGISQFLLGKGVWSPVLKPGVVSVTAYPPRGWFDLFDYTRSVTA 731  
Db 715 RPLVHEFYEDTNSKVEDREFLWGPALLIPLVLTQGAETSAVYIPDAVYDY----- 765  
QY 732 STG-----RYVTLSAPPDHIINVH:QEGNILAMOGKAMTTOARKTPFHLWVMSDCGA 784

Db 766 ETGAKRPRKORVKSIPADKIGLHLAGGYIIPQPAVTTASRMPLGLIILNDONT 825  
QY 785 SPGELFLDDGVETVMGVNRGKMTFVKFAASAKQTCITISDWSGEFAVSQKQVIDKVTI 844  
Db 826 AVGDFFWDDG-ETKQTVQNDQNYILYTFVAVSNLNINLTCTHELSEGTTLA----FOTIKI 880  
QY 845 LGLRKGTYKINGTYTTCANVTRKGDGSKLKS--TPDRKGEFIVAEISGLNLLGREFKL 900  
Db 881 LGVTE-----TVQTVTAENNQSMSTHSNFTYDPSNQVLL--IENLNFNLGRNFRV 929

RESULT 14  
TI0799  
sucrose alpha-glucosidase (EC 3.2.1.48) / oligo-1,6-glucosidase (EC 3.2.1.10) - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000  
C;Accession: TI0799  
R;Chandrasena, G.; Osterholm, D.E.; Sunitha, I.; Henning, S.J.  
Gene 150, 355-360, 1994  
A;Title: Cloning and sequencing of a full-length rat sucrose-isomaltase-encoding cDNA.  
A;Reference number: Z17155; MUID:95121929; PMID:7821806  
A;Accession: TI0799  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-1841 <CHA>  
A;Cross-references: EMBL:L25926; NID:g414818; PIDN:AAA65097.1; PID:g773669  
A;Experimental source: strain Sprague-Dawley, intestine  
C;Genetics:  
A;Gene: SI  
C;Function:  
A;Description: sucrose alpha-glucosidase catalyzes the hydrolysis of the 1,2-alpha-D-  
A;Pathway: carbohydrate digestion  
C;Superfamily: sucrose/isomaltase; sucrose/isomaltase homology; trefoil homology  
C;Keywords: carbohydrate digestion; glycosidase; hydrolase  
F;199-846/Domain: sucrose/isomaltase homology <SIM>  
F;937-983/Domain: trefoil homology <TRF>

Query Match 27.7%; Score 1323; DB 2; Length 1841;  
Best Local Similarity 34.9%; Pred. No. 2.1e-85;  
Matches 298; Conservative 155; Mismatches 280; Indels 122; Gaps 27;

QY 38 YGYKVS-VKDSGTRRSALTALPOLKNSVYGPDIQLLSITASLESNDRLVRITDAKH 96  
Db 122 HGYNASITNENAGLKATLRIP---SPTLFGEDIKSVILTTQTQGNRFRKLTDPNN 177  
QY 97 RWRWELPNILHRHQPPTPPHSLSSLYRTLLSSPTTNRKILLSHPNSDLTFSINTTFF 156  
Db 178 KRYEVPHQVKEETGIP---AADTLVDVQVSE-----NPF 209  
QY 157 GFTISRKSTHVDLPDTPNPNTFLIFIDQYLHLTSSLPGRHAIYGLGHSKPTFQL 216  
Db 210 SIKVIRKSNKVLCDTSVGP-----LLYSNQYLQISTRLPS--EYIYFGGHIHKEFR- 260  
QY 217 AHN---QTLTNRAD-IPSSNPVNLVYSHPPYMDVRSSPVAGSTHGVLLNSNGMDVEY 272  
Db 261 -HDLVMTWPIFTTRDEIPGDN--NENLYGHOTFFMGI--GDTSGKSYGVFLMNSNAMEVFI 316  
QY 273 TGNR-ITKVIIGIIDLFFAGSPGVQVQFTRVIGRPAHPMPYAFQOCRYGHDVY 331  
Db 317 QPTPIITVTVRVIGIIDLFFAGSPGVQVQFTRVIGRPAHPMPYAFQOCRYGHDVY 331  
QY 332 ELQSVVAGYAKAKIPLVWMTDIDYMDAYKDTLPDVPNEPLDKMKKFNVLHKNQKQYV 391  
Db 377 TVSEVVRNRREALIPDQVSDIDTMEKDKQFYDORVAY--NGLPDPFQDLHDHGQKV 433  
QY 392 ILDPGISTNK-----TYETIRGMKHDVFLKRG--KPYLGSVWPGPVYFPDFLKPSALT 444  
Db 434 IILDPAISINRRASGEAYESYDRGNAQVWVNESDGTTPVGEVWPGDTPVDFTPNCI 493  
QY 445 FWTDEIKRFLNLLPVDGLWIDMNEISNFI--SSPIPGSTLDNPPYKKNNGVWLPITNKTI 504  
Db 494 WWANECNLFHOEVNDGLWIDMNEVSSVQSGKNCNDNTLNPY-----SLNLKGVLLITVLN--Y 538



QY 505 PP-----TAMHGDIPYVNVHNLFGYLEARVTRAAIKL-TEKRFV 545  
 DB 539 PPTFGILDKWYSKTLCDVQHWG--KQYDVHSLGYSAIAIEQAVERFPNKSFI 596  
 QY 546 LSRSTFSSGKYTAHTWGNATWDLVYSIPMSLDGLFGIPMWGAICGFLGNTTEEL 605  
 DB 597 LTRSTFGSGRANHLGDNATASWEQMSITGMLFEG-FGMLPGVATSCGFLADTTEEL 656  
 QY 606 CRWTLQCAFYPFSDHSLGTTTQGE--LYRMESVAASARKVGLRYTLIPFYTLMYEAQ 664  
 DB 657 CRWMLQCAFYPFSDHSLGTTTQGE--SSRHYLTIRYTLIPFYTLFYEAH 713  
 QY 665 LMGIPARPLFPFSDPDIDKTYGISSQFLGKGMVSPVLKFGVSVTAFFPRGNWFLDF 724  
 DB 714 MGETVARPLFVFDYDNTSWIEDTQFLGWPALLITPVRQVENVSAYIPNATWYDY-- 771  
 QY 725 YTRSVTSTG-----RYVT-SAPDHNHVIHQENILAMQGMKAMTTOAARKTPFHLV 777  
 DB 772 -----ETGIKRPWRKERINMYLPQDKIGLHLRGYIIPTEPDVTTTASRKNPLGLIV 824  
 QY 778 VMSDCGASGELFLDGVETVGVNRGKMTFVKFAAASAKOTCIITSDVSGEFAVSQW 837  
 DB 825 ALDDNQAKGELFWDG-ESKDSIEKKMILYTFVSNN-----LVLNCTHSSVAEGTSL 879  
 QY 838 VIDKVTILGRKGT 852  
 DB 880 AFKTIKVLGLREDVR 894

RESULT 15  
 UUUU  
 N: Alternate names: limit dextrinase; oligo-1,6-glucosidase (EC 3.2.1.10) [validated]  
 C: Species: Homo sapiens (man)  
 C: Date: 19-Nov-1988 #sequence\_revision 24-May-1996 #text\_change 08-Dec-2000  
 C: Accession: S36082; A27326; S24329; A61136  
 R: Lacasa, M.  
 submitted to the EMBL Data Library, December 1991  
 A: Reference number: S36082  
 A: Accession: S36082  
 A: Molecule type: mRNA  
 A: Residues: 1-1827 <LAC>  
 A: Cross-references: ENBL:X63597; NID:G36644; PIDN:CAA45140.1; PID:G36645  
 R: Green, F.; Edwards, Y.; Hauri, H.P.; Povey, S.; Ho, M.W.; Pinto, M.; Swallow, D.  
 Gene 57, 101-110, 1987  
 A: Title: Isolation of a cDNA probe for a human jejunal brush-border hydrolase, sucrase-1  
 A: Reference number: A27326; MUID:88112852; PMID:2962903  
 A: Accession: A27326  
 A: Molecule type: mRNA  
 A: Residues: 1-661, 'X', 663-678 <GRE>  
 A: Cross-references: GB:M22616  
 R: Chantret, I.; Lacasa, M.; Chevalier, G.; Ruf, J.; Islam, I.; Mantei, N.; Edwards, Y.;  
 Biochem. J. 285, 915-923, 1992  
 A: Title: Sequence of the complete cDNA and the 5' structure of the human sucrase-isomaltase  
 A: Reference number: S24329; MUID:92359963; PMID:1353958  
 A: Accession: S24329  
 A: Molecule type: mRNA  
 A: Residues: 1-661, 'F', 663-931 <CHA>  
 A: Cross-references: ENBL:X63597  
 R: Gorvel, J.P.; Ferrero, A.; Chambraud, L.; Rigal, A.; Bonicel, J.; Marcoux, S.  
 Gastroenterology 101, 618-625, 1991  
 A: Title: Expression of sucrase-isomaltase and dipeptidylpeptidase IV in human small intestine  
 A: Reference number: A61136; MUID:91317403; PMID:1677636  
 A: Accession: A61136  
 A: Molecule type: protein  
 A: Residues: 2-14, 'F', 16-20; 1008-1015, 'E', 1017-1021, 'TX', 1024 <GOR>  
 C: Genetics:  
 A: Gene: GDB:SI  
 A: Cross-references: GDB:120377; OMIM:222900  
 A: Map position: 3q25.2-3q26.2  
 C: Complex: the two product chains remain associated after cleavage  
 C: Function: <ISM>

A: Description: oligo-1,6-glucosidase catalyzes the hydrolysis of 1,6-alpha-D-glucosidi  
 A: Pathway: carbohydrate digestion  
 C: Function: <SUC>  
 A: Description: sucrose alpha-glucosidase catalyzes the hydrolysis of the 1,2-alpha-D-g  
 A: Pathway: carbohydrate digestion  
 C: Superfamily: sucrase/isomaltase; sucrase/isomaltase homology; trefoil homology  
 C: Keywords: carbohydrate digestion; duplication; glycoprotein; glycosidase; hydrolase;  
 F: 2-1007/Product: oligo-1,6-glucosidase (isomaltase chain) #status experimental <MAT1>  
 F: 13-32/Domains: membrane associated #status predicted <TM>  
 F: 42-60/Region: serine/threonine-rich  
 F: 63-109/Domains: trefoil homology <TRF1>  
 F: 189-840/Domains: sucrase/isomaltase homology <SIM>  
 F: 931-977/Domains: trefoil homology <TRF2>  
 F: 1008-1827/Product: sucrose alpha-glucosidase (sucrase chain) #status experimental <M  
 F: 1062-1734/Domains: sucrase/isomaltase homology <SIM2>  
 F: 12/Binding site: carbohydrate (Ser) (covalent) #status absent  
 F: 99,437,455,823,855,904,926,1235,1303,1340,1535,1748,1763,1815/B1  
 F: 1007-1008/Cleavage site: Arg-Ile (cysteine) #status predicted

Query Match 27.7% Score 1318.5 DB 1; Length 1827;  
 Best Local Similarity 35.1%; Pred. No. 4.3e-85;  
 Matches 319; Conservative 162; Mismatches 298; Indels 129; Gaps 32;  
 QY 35 VIGYGVKSVKVDG--GTRRSITLALPQLVKNSSVYGPDIQLLSITASLESNDRLVRITD 93  
 DB 109 VDNHGVNQDMTTTISIGVEAKLRIP-----SPTLFGNDINSVLFTTQNTQTPNRFKIID 164  
 QY 94 AKHRWEIPDNILHRHQPPPPPHSSUSSYRTLLSPTNRRKILLSHNSDUTFSL-IN 152  
 DB 165 PNNRYEVP-----HQ-----YVKEFTGTV-----SDTLVDVKA 195  
 QY 153 TTPFGFTISKSTHVDLDPDTPNPNTFLIPIDOLYHLTSSLPGTRAHIVGLGS--HS 210  
 DB 196 QNPFSTQVIRKNGKTLFTDSIGP-----LVSDQYLQISARLPSD--YIYGIGSQVHK 247  
 QY 211 KPTFQLAHNQTLTMRADIPSSNPVNLXGSHFVYNDVSSPVAGSTHGVLLNSNGMDV 270  
 DB 248 RFRHDLWSKTPITRDQLPGDNN--NLXGHQTFNCIEDT--SGKSGFVFLMNSNAMEI 304  
 QY 271 EYTGRI--TYKVJGGIIDLFFAGPSFGVQVQFTVIGRPAPMPYAFGQOCRYGHD 329  
 DB 305 FIQPTPIVTVYVTGGILDFYILLGDTPEOVQOQLGLVPLAMPAYMNLGFSQLSRNYS 364  
 QY 330 VYELQSVAGYAKAKIPLVMTMDIDYMDAYKDFIDVNFPLDKMKKFNVLNKHGQKY 389  
 DB 365 LDVKEVVRNREAGIPFDQVTDIDYMDKDKFTYDQVAF--NGLPQVQDLHDHGQKY 422  
 QY 390 VILDFGISTNK-----TYETIRGMKHDVFLKRN--GRPYLGSVMPGVYFPDFLKP 442  
 DB 423 VILDPALISGRANGTTVATYERGNTQHWINESDGSPTIIGVWPGLTVYDPFTNPN 482  
 QY 443 LTFWTDIELKFLNLLPVDGLWDMNEISNIPSPPIPGST-----LNPPIKINNSGV 495  
 DB 483 IDWANECSIFHOVQYQVGLWDMNEVSSFIQ-----GSTKGNVKNLNYPPPTDILDK 537  
 QY 496 MLPIINKTPTTAM--HYGDIPEYVNVNLFGLYEAARVTRAAIKL--TEKRPVLSRSTFS 553  
 DB 538 LM--YSKTIKMDAVQWNG--KQYDVHSLGYSAIAIEQAVERFPNKSFI 593  
 QY 554 SGKYTAHTWGNATWDLVYSIPMSLDGLFGIPMWGADICGFLGNTTEELCRRWIQLG 613  
 DB 594 SGRHAHLGDNATASWEQMSITGMLFEGSLFGIPLVGADICGFVAETTEELCRRWIQLG 653  
 QY 614 AFYFPRDHSLLGTTTQGE--LYRMES--VAASARKVGLRYTLIPFYTLMYEAQINCIPI 670  
 DB 654 AFYFPRNHSNDSYEHQDPAFFQNSLLVKSQRQYLTIRYTLIPFYTLFYKAHVFETV 713  
 QY 671 ARPLFFSFPDDIKTYGISQFLGKGMVSPVLKFGVSVTAFFPRGNWFLDFYTRSVT 730  
 DB 714 ARPVLEHFEVDNTSWIEDTEFLWGPALLITPVLKQAGDVTSAIYIPDAIWDY----- 765  
 QY 731 ASTG-----RYVTLSAPPDHNHVIHQENILAMQGMKAMTTOAARKTPFHLVMSDCG 783



Db	766	-ESGAKPRKRVNDYLPADKIGLHRLGGYIIPIQEPDVTTTASRNKPLGLI VALGNN	824
Qy	784	ASGELFLDGGVEVTMGVNRGKWTFFVKIAASAKQTCIITSDVSGEFASQKWWIDKVT	843
Db	825	TAKGDFFWDDG-ETKDTIQNGYIIYTSVNN-----TLDIV-----CT	863
Qy	844	ILGRKGT KINGTVR---TGAVT--RKGDKSKLKS-----TPDRKGEFIVAEISGLNL	892
Db	864	HSSYQEGTTLAFTQVKILGTLDSVEVRVAENQPMNAHSNFTYDASNOVLL--IADLKL	921
Qy	893	LLGREPKL 900	
Db	922	NLGRNFSV 929	

Search completed: October 27, 2003, 10:22:10  
Job time : 27.6046 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:20:43 ; Search time 108.481 Seconds  
(without alignments)  
1393.955 Million cell updates/sec

Title: US-10-043-418-3  
Perfect score: 4768  
Sequence: 1 MKKKIPSLALGILLVFLQY.....VAEISGLNLLGREFKLVJH 903

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 629382 seqs, 167460530 residues

Total number of hits satisfying chosen parameters: 629382

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US05\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09D\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	4768	100.0	903	14	US-10-043-418-3	Sequence 3, Appli
2	3243.5	68.0	914	14	US-10-043-418-2	Sequence 2, Appli
3	2830.5	59.4	902	14	US-10-043-418-4	Sequence 4, Appli
4	2422.5	50.8	877	14	US-10-043-418-1	Sequence 1, Appli
5	1880	39.4	707	12	US-10-032-189-125	Sequence 125, App
6	1318.5	27.7	1827	9	US-09-819-247-2	Sequence 2, Appli
7	851	17.8	914	12	US-10-032-189-42	Sequence 42, Appl
8	848	17.8	914	12	US-10-032-189-40	Sequence 40, Appl
9	847	17.8	912	12	US-10-032-189-44	Sequence 44, Appl
10	805.5	16.9	943	12	US-10-032-189-123	Sequence 123, App
11	805.5	16.9	944	12	US-10-032-189-122	Sequence 122, App
12	805	16.9	966	12	US-10-032-189-121	Sequence 121, App
13	804	16.9	966	12	US-10-032-189-120	Sequence 120, App
14	799	16.8	944	12	US-10-032-189-124	Sequence 124, App
15	797.5	16.7	967	12	US-10-032-189-38	Sequence 38, Appl

16	715.5	15.0	565	15	US-10-102-806-557	Sequence 557, App
17	681.5	14.3	718	12	US-10-228-063-26	Sequence 26, Appl
18	681.5	14.3	718	12	US-10-228-063-36	Sequence 36, Appl
19	676	14.2	693	12	US-10-228-063-5	Sequence 5, Appli
20	676	14.2	712	12	US-10-228-063-27	Sequence 27, Appl
21	590.5	12.4	235	9	US-09-734-569-132	Sequence 132, App
22	588.5	12.3	788	15	US-10-156-761-14497	Sequence 14497, A
23	487	10.2	1070	10	US-09-280-197-6	Sequence 6, Appli
24	487	10.2	1070	10	US-09-423-126-4	Sequence 4, Appli
25	476	10.0	1066	10	US-09-280-197-5	Sequence 5, Appli
26	476	10.0	1066	10	US-09-423-126-3	Sequence 3, Appli
27	441	9.2	642	15	US-10-156-761-10107	Sequence 10107, A
28	393	8.2	188	9	US-09-734-569-134	Sequence 134, App
29	371.5	7.8	1092	10	US-09-423-126-5	Sequence 5, Appli
30	355.5	7.5	1088	10	US-09-280-197-1	Sequence 1, Appli
31	355.5	7.5	1088	10	US-09-423-126-1	Sequence 1, Appli
32	350	7.3	1091	10	US-09-280-197-2	Sequence 2, Appli
33	350	7.3	1091	10	US-09-423-126-2	Sequence 2, Appli
34	335	7.0	570	10	US-09-423-126-6	Sequence 6, Appli
35	306	6.4	199	9	US-09-734-569-130	Sequence 130, App
36	129	2.7	317	10	US-09-280-197-9	Sequence 9, Appli
37	124.5	2.6	2328	15	US-10-171-311-64	Sequence 64, Appl
38	124.5	2.6	2386	11	US-09-961-403-1	Sequence 1, Appli
39	122.5	2.6	2320	12	US-10-279-733-8	Sequence 8, Appli
40	120	2.5	3354	12	US-10-174-677-9	Sequence 9, Appli
41	120	2.5	3354	15	US-10-160-758-11	Sequence 11, Appl
42	120	2.5	3354	15	US-10-160-758-12	Sequence 12, Appl
43	114	2.4	740	15	US-10-155-400-3	Sequence 3, Appli
44	114	2.4	740	15	US-10-155-400-6	Sequence 6, Appli
45	114	2.4	957	15	US-10-155-400-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-10-043-418-3  
; Sequence 3, Application US/10043418  
; Publication No. US20020184662A1  
; GENERAL INFORMATION:  
; APPLICANT: Henson, Cynthia A.  
; APPLICANT: Muslin, Elizabeth H.  
; APPLICANT: Clark, Suzanne E.  
; TITLE OF INVENTION: Modified barley alpha-glucosidase  
; FILE REFERENCE: 960296.97486  
; CURRENT APPLICATION NUMBER: US/10/043,418  
; CURRENT FILING DATE: 2002-06-25  
; PRIOR APPLICATION NUMBER: 60/260,787  
; PRIOR FILING DATE: 2001-01-10  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 903  
; TYPE: PRT  
; ORGANISM: Spinach  
US-10-043-418-3

Query Match	100.0%	Score 4768;	DB 14;	Length 903;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 903;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Cy	1	MKKKIPSLALGILLVFLQYLVAGISTSENDEPGVIGYKVKSVKVDSTRRSLTALPQ	60	
Db	1	MKKKIPSLALGILLVFLQYLVAGISTSENDEPGVIGYKVKSVKVDSTRRSLTALPQ	60	
Oy	61	LVKNSSVYGPDIIQLLSITASLESNDRLRVITDAKHRRWEIPDNILHRHQPDPHPSLS	120	
Db	61	LVKNSSVYGPDIIQLLSITASLESNDRLRVITDAKHRRWEIPDNILHRHQPDPHPSLS	120	
Oy	121	SLVRTLSSPTTNRKILLSHPNSDLTFSLINTTTPGFTISRKSTHDVLFDATPDPTNPN	180	
Db	121	SLVRTLSSPTTNRKILLSHPNSDLTFSLINTTTPGFTISRKSTHDVLFDATPDPTNPN	180	

181 TFLIFIDQYLHLTSSLPGRTRAHYGLGHSKPTFQLAHNQTLMRAADIPSSNPVDNLVG 240  
Db |||||  
181 TFLIFIDQYLHLTSSLPGRTRAHYGLGHSKPTFQLAHNQTLMRAADIPSSNPVDNLVG 240  
Qy |||||  
241 SHPFYMDVRSSPVAGSTHGVLLNSNGMDVEYTGNRITYKVGIGIIDLYFFAGSPGQVV 300  
Db |||||  
241 SHPFYMDVRSSPVAGSTHGVLLNSNGMDVEYTGNRITYKVGIGIIDLYFFAGSPGQVV 300  
Qy |||||  
301 EQTRVIGRAPMPYAFQOCRYGHDVYELQSVVAGYAKAKIPLVMMWTDIDYMDAY 360  
Db |||||  
301 EQTRVIGRAPMPYAFQOCRYGHDVYELQSVVAGYAKAKIPLVMMWTDIDYMDAY 360  
Qy |||||  
361 KDFLDPVNFPLDKMKFVNNLHNGQKYVVIDPGISTNKTYETVIRGMKHDVFLKRG 420  
Db |||||  
361 KDFLDPVNFPLDKMKFVNNLHNGQKYVVIDPGISTNKTYETVIRGMKHDVFLKRG 420  
Qy |||||  
421 KPYLGSVWPGVYFPDFLKPALSALTFTWDEIKRFLNLLPVDGLWIDMNEISNFISSPIPG 480  
Db |||||  
421 KPYLGSVWPGVYFPDFLKPALSALTFTWDEIKRFLNLLPVDGLWIDMNEISNFISSPIPG 480  
Qy |||||  
481 STLDNPPYKINNSGVMLPIINKTIPTTAMHYGDIPEYVNHNLPGYLEARVTRAALIKLTE 540  
Db |||||  
481 STLDNPPYKINNSGVMLPIINKTIPTTAMHYGDIPEYVNHNLPGYLEARVTRAALIKLTE 540  
Qy |||||  
541 KRPFVLSRSTFGSGKYTAHWTGDNAAATWMDLVYSIPSMDFGLFGIPMVGADICGFLGN 600  
Db |||||  
541 KRPFVLSRSTFGSGKYTAHWTGDNAAATWMDLVYSIPSMDFGLFGIPMVGADICGFLGN 600  
Qy |||||  
601 TTTELCCRWTQLGAFYFPDRDSSLTGTYOELYRWESVAASARKVGLRYTLPPYFTLM 660  
Db |||||  
601 TTTELCCRWTQLGAFYFPDRDSSLTGTYOELYRWESVAASARKVGLRYTLPPYFTLM 660  
Qy |||||  
661 YEALQNGIPIARPLPFSFPDDIKTYGISSOFLGKGVMVSPVLKPGVSVTAYFPRGNWF 720  
Db |||||  
661 YEALQNGIPIARPLPFSFPDDIKTYGISSOFLGKGVMVSPVLKPGVSVTAYFPRGNWF 720  
Qy |||||  
721 DLFDYTRSVTASTGRYVTLSSAPDHNHVIQEGNILAMQKAMTTQAARKTPHLLVWMS 780  
Db |||||  
721 DLFDYTRSVTASTGRYVTLSSAPDHNHVIQEGNILAMQKAMTTQAARKTPHLLVWMS 780  
Qy |||||  
781 DCGASFGELFDGVEVTMGVNRGKWTFFVKFIAASAKOTCIITSDVVSSEFAVSQKWID 840  
Db |||||  
781 DCGASFGELFDGVEVTMGVNRGKWTFFVKFIAASAKOTCIITSDVVSSEFAVSQKWID 840  
Qy |||||  
841 KVTILGLRKGTNGYTVRTGAVTRKGDKSKLSTPDRKGEFVAEISGLNLLGREFKL 900  
Db |||||  
841 KVTILGLRKGTNGYTVRTGAVTRKGDKSKLSTPDRKGEFVAEISGLNLLGREFKL 900  
Qy VLH 903  
Db VLH 903

## RESULT 2

US-10-043-418-2  
; Sequence 2, Application US/10043418  
; Publication No. US20020184662A1  
; GENERAL INFORMATION:  
; APPLICANT: Henson, Cynthia A.  
; APPLICANT: Muslin, Elizabeth H.  
; APPLICANT: Clark, Suzanne E.  
; TITLE OF INVENTION: Modified barley alpha-glucosidase  
; FILE REFERENCE: 960296.97486  
; CURRENT APPLICATION NUMBER: US/10/043.418  
; PRIOR FILING DATE: 2002-06-25  
; PRIOR FILING DATE: 2001-01-10  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 914  
; TYPE: PRF  
; ORGANISM: Sugar beet

US-10-043-418-2

Query Match 68.0%; Score 3243.5; DB 14; Length 914;  
Best Local Similarity 67.3%; Pred. No. 1.8e-292;  
Matches 610; Conservative 122; Mismatches 166; Indels 9; Gaps 5;

Qy 2 KKKIPSL--ALGILLVFLQLVAGISTSENDPEG-VIGYGVKVKVVDGSTRRLSLTA 57  
Db 3 RSKLPRIYCPTLAVVLPLVLCWVEGATTSKQNDQGEAIGYGVQVNAKVDNSTGKSLTA 62  
Qy 58 LPQLVKNSSVYGPDIQLLSITASLESNDRLRVITDAKRRRWEIPDNILHRHQPPPPPH 117  
Db 63 LLQIRNSPVYGPDIHFLSFTASFEEDTLRIRPTDANRRRWEIPEVLPR-ppppsp 121  
Qy 118 SLSSLYRTLSSPTNRKILLSHPNSDLTFLSINTTPTFGFTISRKSTHVDLFDATDPT 177  
Db 122 PLSSLOHLPKPIPONQPTTVLSSHPSDLAFTLFHTTFFGFTIYRKSTHVDLFDATDPT 181  
Qy 178 NPNTFLIFIDQYLHLTSSLPGRTRAHYGLGHSKPTFQLAHNQTLMRAADIPSSNPVDN 237  
Db 182 NPTTFLIVKQYQLQSSSLPAQQAHLXGLGHTKPTFQLAHNQTLMNADIASFNRDLN 241  
Qy 238 LYGSHPFYMDVRSSPVAGSTHGVLLNSNGMDVEYTGNRITYKVGIGIIDLYFFAGSPG 297  
Db 242 LYGSHPFYMDVRSSPVAGSTHGVLLNSNGMDVEYTGDRITYKVGIGIIDLYIFAGTRPE 301  
Qy 298 QVVSQFTRVIGRAPMPYAFQOCRYGHDVYELQSVVAGYAKAKIPLVMMWTDIDYM 357  
Db 302 MVLDQYTKLIGRAPMPYAFQOCRYGHDVYELQSVVAGYAKAKIPLVMMWTDIDYM 361  
Qy 358 DAYKDFLDPVNFPLDKMKFVNNLHNGQKYVVIDPGISTNKTYETVIRGMKHDVFLK 417  
Db 362 DAFKDFLDPVNFPLDKMKQVFTKLRNGQRYVPLDPIGINTKSYGTFIRGMQSNVFIK 421  
Qy 418 RNGKPYLGSVWPGVYFPDFLKPALSALTFTWDEIKRFLNLLPVDGLWIDMNEISNFISSPP 477  
Db 422 RGNPYLGSVWPGVYFPDFLDPAAFSFVDEIKRFRDILPIDGIMWDMREASNFITSA 481  
Qy 478 IPGSTLDNPPYKINNSGVMLPIINKTIPTTAMHYGDIPEYVNHNLPGYLEARVTRAALIK 537  
Db 482 TPGSTLDNPPYKINNSGVRVINSKTIPTATAMHYGNVTEYNAHNLGFLSQATREALVR 541  
Qy 538 LTKRPFVLSRSTFGSGKYTAHWTGDNAAATWMDLVYSIPSMDFGLFGIPMVGADICGF 597  
Db 542 PATRGPFLLSRSTFAGSGKYTAHWTGDNAAARWDDLOYSIPTMLAFGLFGMPMICADICGF 601  
Qy 598 LONTTEELCCRMIQLGAFYFPDRDSSLTGTYOELYRWESVAASARKVGLRYTLPPYFY 657  
Db 602 AESTTEELCCRMIQLGAFYFPDRDSSLTGTYOELYRWESVAASARTVLGLRYELLPPY 661  
Qy 658 TLMYEAQLNGIPIARPLPFSFPDDIKTYGISSOFLGKGVMVSPVLKPGVSVTAYFPRG 717  
Db 662 TLMYDANLRGPIARPLSFTFPDDVATYIGISSOFLGIRGIMVSPVLQGGSSIVNAYSPRG 721  
Qy 718 NMFDLFDYTRSVTASTGRYVTLSSAPDHNHVIQEGNILAMQKAMTTQAARKTPHLLV 777  
Db 722 NWVLSLNTYSSVSAGTYVSLSAPPDHNHVIHEGNIIVAMQGEAMTTQAARSTPFHLLV 781  
Qy 778 VMSDCGASFGELFDGVEVTMGVNRGKWTFFVKFIAASAKOTCIITSDVVSSEFAVSQKW 837  
Db 782 VMSDHVASTGELFLDNGIEMDIPGGKWTILVRFAESGINNLTISSEVNRGVAMSQRW 841  
Qy 838 VIDKVTILGLRKGTNGYTVR--TGAVTRKGDKSKLSTPDRKGEFVAEISGLNLLG 895  
Db 842 VNDKLTILGLRKVKIKEYTVQKDAIKVKGGLGRRTSS--HNQGGFFVSVISDLRQLVG 899  
Qy REFLEVL 902  
Db QAFKLEL 906

RESULT 3  
US-10-043-418-4

```
; Sequence 4, Application US/10043418
; Publication No. US20020184662A1
; GENERAL INFORMATION:
; APPLICANT: Henson, Cynthia A.
; APPLICANT: Muslim, Elizabeth H.
; TITLE OF INVENTION: Modified barley alpha-glucosidase
; FILE REFERENCE: 960296.97486
; CURRENT APPLICATION NUMBER: US/10/043,418
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: 60/260,787
; PRIOR FILING DATE: 2001-01-10
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 902
; TYPE: PRT
; ORGANISM: Arabidopsis
; US-10-043-418-4

Query Match      59.4%; Score 2830.5; DB 14; Length 902;
Best Local Similarity 59.6%; Pred. No. 51e-254;
Matches 540; Conservative 138; Mismatches 205; Indels 23; Gaps 8;

QY 6 PSALGILLVFLQYLIVAGISTSENDEGVGYKVKSVKVDGSTRSLTALPOLVQNS 65
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 8 PNIFVWVFFSLR---SSQWLEEBESTWGYGVVRSVGVD-SNRQVLTAKLDLXPS 63
QY 66 SVYGPDIQLLSITASLESNDRLVRITDAKRRWEIPDNILHR--HQPPPPHSLSL 122
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 64 SVYAPDLSLNHVSLETSLRIRITDSQQRWEIPETVIPRAGNHSP-----RRFST 117
QY 123 YRTLLSPPTNRKILLSHPNSDLTFTSLINTTFFGTISRKSTHDLVLPDTPNPTNF 182
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 118 EEDGNSPENN---FLADPSSDLVTLHTTFFGSVSRSSGDLFTSDSDSNTY 173
QY 183 LIFIDQVLHUTSSLPGTRAHIYGLGSHSKPTFQLAHNQTLTWRADIPSSNDVNLVYGH 242
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 174 FIFKDFQLQSSALPENRSLYIGISHTKRSFRLIPGETMTLWNADIGSENPDVNLVYGH 233
QY 243 PFYMDVRSS---FVAGSTHGVLLNSGMDVEYTGRIYKVGIGIDLYFFAGSPGV 299
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 234 PFYMDVRSGKNEEAGTHGVLLNSGMDVYEGHRIYTNVIGVIDLYVFGSPGV 293
QY 300 VEOFTVRIGRPAMPYMAFGQCRGYGHDVYELQSVVAGYAKAKIPILEVMTDIDYMDA 359
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 294 MNQYTELIGRPAMPYMSFGHCRGYKKNVSDLEYVVDGYAKAG:PLEVMTDIDYMDG 353
QY 360 YKDFTLDPVNFPLDKMKKFVNNLHKQKQYVWLDLFGISTNKTYETIIRGMKHDVFLQKN 419
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 354 YKDFTLDPVNFPEDKMQSFYDTLHKQKQYVWLDLFGIGVDDSYGTYNRMGEADVIFKRN 413
QY 420 GKPYLGSVMPGPVYFPDFLKPXSALTFTWDEIKRFLNLLPVDGLWIDMNEISNFISSPP 479
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 414 GEYPLGEVWPKYVFPDFLNPAAATFWSNEIKMFQELPLDGLWIDMNEISNFISSPLS 473
QY 480 GSTLDNPPYKINKSGVWMLPIINKTPIPTAMHYGDIPYVNN--FGYLEARVTRAALIKIT 539
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 474 GSSLDPPPYKINKSGDKRPINNKTVTPATSTHFCNISEYDAHNLVLEAKATHQAVVDIT 533
QY 540 EKBPFLVSRSTFSGSKYTAHTGDNAATNDLVYIPSMDFLFGIPMWGADICGFLG 599
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 534 GKRPFLVSRSTFVSSGKYTAHTGDNAAKWEDLAYSIPG:LNFLGFIPIWVGADICGFSH 593
QY 600 NTEELCRRIQIOLGAFYPPSRDSSLGTTQELYRWESVAASARKVLGLRYT:LPIVEYTL 659
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 594 DTTEELCRRIQIOLGAFYPPARDSSLGTAQELYLWDSVASSARKVLGLRMILLPLHYTL 653
QY 660 MYEQLNGPIARPLFPFSDDKTYGISSQFLGKGVWVSPVLKPGVSVTAYFPRGNW 719
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 654 MYEAVHSGNPIARPLFPFSDDKTYEIDSQFLGKSIWSPALKQGAVALDAYFPAGNW 713
QY 720 FDLFDYTRSVTASTGRVTVLSAPDPDHNVHIQSGNILLMOGKAMTQAARKTFPHLLVVM 779
```

```
114 FDLENYSFVAVGDSGKGVRLDTADHVNHVHREGSI VAMQGEALTTRDARKTPVQLLWA 773
QY 780 SDCGASFGELFLDGVVVTMGVNRCK--WTFVVFIAASAKQTCIITSDDVVSFGFVSKW 837
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 774 SRLENISGELFLDDGLENRMGAGGGRDWTLVKPRCYVTVGKSVVLRSEVVPFAYASKW 833
QY 838 VIDKVTILGRKGTKINGYTVRTGAVTRKDGKSKLKSTPDRKG-EFIVAEISGLNLLGR 896
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 834 SIGKVTTFGVENVENKTYEVRTSERLSRISLIKTIVSNDOPRFLSVKSLVGVK 893
QY 897 EFKLVL 902
Dd : : : : :
Dd 894 KFMRL 899

RESULT 4
US-10-043-418-1
; Sequence 1, Application US/10043418
; Publication No. US20020184662A1
; GENERAL INFORMATION:
; APPLICANT: Henson, Cynthia A.
; APPLICANT: Muslim, Elizabeth H.
; APPLICANT: Clark, Suzanne E.
; TITLE OF INVENTION: Modified barley alpha-glucosidase
; FILE REFERENCE: 960296.97486
; CURRENT APPLICATION NUMBER: US/10/043,418
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: 60/260,787
; PRIOR FILING DATE: 2001-01-10
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 877
; TYPE: PRT
; ORGANISM: Barley
; US-10-043-418-1

Query Match      50.8%; Score 2422.5; DB 14; Length 877;
Best Local Similarity 53.6%; Pred. No. 4.9e-216;
Matches 486; Conservative 132; Mismatches 234; Indels 55; Gaps 15;

QY 10 LGILLVFLQYLIVAG--ISTSENDP---EGVIGYGVKVK-SVKVDST-RSLTALPOLV 62
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 4 VGVLLCLCLCLPAPRLCSSKEGFLAARTVLA VAVTMEGALRAEAATGCRSSTG----- 58
QY 63 KNSVGYPDQLLSITASLESNDRLVRITDAKRRWEIPDNILHRHOP-----PPPPHS 118
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 59 -----DVORLAVYASLETDSRLVRITDAHPRWEVQDIIIPRAPGDVLHDAPAS 110
QY 119 LSSLYRTLSSPTNRKILLSHPNSDLTFTSLINTTFFGTISRKSTHDLVLPDTPN 178
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 111 -----SAPLQGR---VLSPAGSDLVLT-VHASPFRTVSRRTGDTLFDTPAG--- 154
QY 179 PNTFLIFIDQVLHUTSSLPGTRAHIYGLGSHSKPTFQLAHNQTLTWRADIPSSNDVNL 238
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 155 ----LVPRDKYLEVTSALPAGRASLYGLGHTKSSFRLRHNDSTFLWNADIGASYVDVNL 210
QY 239 YGSHPFYMDVRSSPVAGSTHGVLLNSGMDVEYTGRIYKVGIGIDLYFFAGSPGV 298
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 211 YGSHPFYMDVRA---PGTAHGVLLSSNGMDVLYGGSYVTVYKVGIVLDFYFAGFAPLA 267
QY 299 VVEQFTRVIGRPAMPYMAFGQCRGYGHDVYELQSVVAGYAKAKIPILEVMTDIDYMD 358
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 268 VVDQYTLIARPAAMPYMSFGHCRGYLNSDLERVARVAKARIPLEVMTDIDYMD 327
QY 359 AYKQFTLDPVNFPLDKMKKFVNNLHKQKQYVWLDLFGIST--NKTYYIIRGMKHDV 415
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 328 GFKDFTLDRVNFATAELRPFVDRLHRNAQKYVWLDLFGIRVDPIDATYGTFRVGMQDIF 387
QY 416 LKRNKQVPLGSVMPGPVYFPDFLKPXSALTFTWDEIKRFLNLLPVDGLWIDMNEISNFI 475
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 388 LKRNKGTVMGVNPGDVFPPDFMHPAAAEFWAREISLFRRTIPVDGLWIDMNEISNFI 446
```

```

Qy 476 PPIPCSTLDNPPYKINNSGVMLPIINKTIPPTAMHYGDIPEYVNVNLFGLVLEAVTRBAAL 535
Db 447 -PEPNALDDPPYINNDGTGRPINNKTVPLAVHYGVGYEYEHNLFGLEARATGRGV 505
Qy 536 IKLTEKRPVLSRSTFSGSGKYTAHWTGDNAATWNDLVYISIPSLDFGLFGIPMVGADIC 595
Db 506 LRDTRRPVLSRSTFVSGSGRYTAYTGTCDNAATWGLDRLYSINTWLSFGLFMPMIGADIC 565
Qy 596 GFLGNTEELCRRWTOIGAFYFSDHSSLGTTTQELRYHESVAASARKVLGLRYTLIPY 655
Db 566 GFNGNTEELCGRWIOIGAFYFSDHSSAIFTVRELYLWPSVAASGRKALGLRYQLPY 625
Qy 656 FYTLMYEAQLNGIPIARLPFSFFDDIKTYGISQFLLGKGVMSVPLKPGVSVTAAYFP 715
Db 626 FYTLMYEAHMTGAPLARLPFSYFHDVATYGVDRQFLLGRGVLSVPLEPGPTIVDAYFP 695
Qy 716 RGNWFDLFDYTRSVTASTGRVYVTLISAPPDHINKVHIQEGNILAMOGKAMTTOAKRTFHL 775
Db 686 AGRWYRLYDYSLAVATRTGKHVLPAPADTVNVHLTGCTILPLQOSALTTSRARRTAHL 745
Qy 776 LVNMSDCGASFGELFLDDGVETWGVNRGKWTFKYF--IAASAKQTCIITSDVVUSGEFV 833
Db 746 LVALAEOGTASGYLFLDDGUSPEYG--RRSDMSWVRFNKYIPNNKGAIKVSEVVRHSIAQ 804
Qy 834 SQKVVIDKVTILGRKGTCKINGYVTRTGAVTRKGDKSKLKSIPDRKGFEFVABISGLNLL 893
Db 805 SRTLVIKVVJMGHRSPPAAPKLLTVHNSAEVEASSAGTRYQVAGGLGVVAHIGGLSLV 864
Qy 894 LGREFKL 900
Db 865 VGEFEL 871

```

RESULT 5

```

US-10-032-189-125
; Sequence 125, Application US/10032189
; Publication No. US20030170630A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook II, John P
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Patturajan, Meera
; APPLICANT: Grosse, William M
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Shimkets, Richard A
; APPLICANT: Grosse, William M
; APPLICANT: Szekeres, Edward S
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Li, Li
; APPLICANT: Casman, Stacie C
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Gorman, Linda
; APPLICANT: Gangolli, Esha A
; APPLICANT: Fernandes, Elma R
; APPLICANT: Rieger, Daniel K
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Gunther, Erik
; APPLICANT: Millet, Isabelle
; APPLICANT: Sciore, Paul
; APPLICANT: Ellerman, Karen
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glennda
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-228
; CURRENT APPLICATION NUMBER: US/10/032,189
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,495
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/258,171

```

```

; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/269,940
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 60/274,192
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/277,826
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/279,840
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/282,981
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/283,656
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/309,247
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/311,754
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/313,331
; PRIOR FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 260
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 125
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Glycosyl
; OTHER INFORMATION: hydrolases family 31
US-10-032-189-125

```

```

Query Match 39.4%; Score 1880; DB 12; Length 707;
Best Local Similarity 53.7%; Pred. No. 1.1e-165;
Matches 383; Conservative 97; Mismatches 208; Indels 28; Gaps 14;

Qy 140 SHPNSDLTSLINTPPGFTTSRKSSTHVDLFDATPDPPNMTFLIFIDQYHLHTSSLPGT 199
Db 9 SPAESDLYDLVNSNGPFGFVEIRKSTGVDLFTTGP-----LVFSQFLQLSTSLPS- 61
Qy 200 RAHYGLGHEHSHKPTFLQAHNOTLTMRADIPSSNPDVNLGSHPEYMDVRSPVAGSTHG 259
Db 62 -EYIYGLGHEHAHLPRRDTNETYTLWNRDVGYPYSGDNNLYGSHHPYMSLEDS---GNAHG 117
Qy 260 VLLNSNGMDVEY-TGNRITYKVGIIIDLYFFAFSPGVQVEQTRVIGRPAPMPYNAP 318
Db 118 VFLNSNAMEVDIGCPALTYRVIGGILDYFFLGPTPEDVLQVYTELIGRALPPYNSL 177
Qy 319 GFQOCRYGHDVYELQSVVAGYAKAKIPLVWMTDIDMDAYKDTLDPVNFPLDKMKKF 378
Db 178 GFHLCRWGYTVVSEVKTVDGKRKANIPLDVQWLDIDYMDGYKDTWDPVRFP--GPEDF 235
Qy 379 VNNLHKGQKYVWILDGIGSTNK-TYETYIRGMKHDVFLKR-NGKPYLGSVMPGVVFPD 436
Db 236 VKLHAKQKYVWILDPAISVDSASYPYERKEGKGVFKVKNPNSDYIGEVWPGVTAEPD 295
Qy 437 FLKPSALTFTWDEIKRFLNLLPVDGLWIDMNEISNFISPPIPGSTLDNPKYKINNSGV 496
Db 296 FNPPEARWMADEIKDFHDSLFPDGIWIDMNEPSP-SEPGPNSNLNYPYAPNDGDG- 353
Qy 497 LPINKTIPPTAMHYGDIPEYVNVNLFGLVLEAVTRAAALIKLT-EKRPVLSRSTFSGSG 555
Db 354 -PLSSKTMCDAVHYGVEYHDVHNLVGLSEAKATYEALKKVTGKRPVLSRSTFAAGS 412
Qy 556 KYTAHWTGDNAATWNDLVYISIPSLDFGLFGIPMVGADIICGFLGNTTEELCRRWTOIGAF 615
Db 413 RYAGHWTGDNATASMDLKYISIPGVLSFNLFGIPFVGADICGFGNGTTEELCRRWTOIGAF 472
Qy 616 YPFSRDHSSLGTTTQELRYHESVAASA-RKVLGLRYTLIPYFYLMYEAQLNGIPIARLP 674
Db 473 YPFSRNHNLGTIQEPLFDSVAASRKALNRYTLPLPYLYTLFHEAHVSGLPVNRPL 532
Qy 675 PFSFPDDIKTYGISQFLLGKGVMSVPLKPGVSVTAAYFPRGNWFDLFDYTRSVTASTG 734
Db 533 PFEFPDDAARTYDIDRQFLWGSALLVAPVLEPGATSVKAYLPGGRWYDL--YTGAGEASRG 590

```

```

QY      735 RVTLSAPPDHINVIHQGNILAMCGKAMITTOAARKTPFHLLVUMSDCASFGELFLDDG   794
        ||||| : |::| ::|::| : ||||| : |::|::| : |::|::| : |::|::|
Db      591 GNVTL$APLDKI$PVHVGRGS$IPTQEPAI$TTESDRNPFHL$VALDCDNGTASGELYLDDG   650
QY      795 VETVMGVNRGKWTFVKFIAASKATCII$TSDW$GEFAVSOKWVIDKVTLGL    847
        ::|::| : |::| : |::| : |::| : |::| : |::| : |::| :
Cb      651 ESI--DTQRGDVLLVQF---SANNNTLTGTVEVTGYKYKNSTNLTKLEKITLV    698

RESULT 6
US-09-819-247-2
; Sequence 2, Application US/09819247
; Patent No. US20010036635A1
; GENERAL INFORMATION:
; APPLICANT: Waldman, Scott A.
; APPLICANT: Park, Jason
; APPLICANT: Schulz, Stephanie
; TITLE OF INVENTION: Compositions And Methods For Identifying And Targeting (c)
; TITLE OF INVENTION: Alimentary Canal Origin
; FILE REFERENCE: TJU2413
; CURRENT APPLICATION NUMBER: US/09/819,247
; CURRENT FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/192,229
; PRIOR FILING DATE: 2000-03-27
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 1827
; TYPE: PR1
; ORGANISM: Homo sapiens
US-09-819-247-2
```

## RESULT 7

US-10-032-189-42

: Sequence 42, Application US/10032189

Publication No. US20030170630A1

GENERAL INFORMATION:

APPLICANT: Alsobrook II, John P

APPLICANT: Tchernev, Velizar T

APPLICANT: Liu, Xiaohong

APPLICANT: Spytek, Kimberly A

APPLICANT: Zerhusen, Bryan D

APPLICANT: Patturajan, Meera

APPLICANT: Grosse, William M

APPLICANT: Lepley, Denise M

APPLICANT: Burgess, Catherine

; APPLICANT: Shimkets, Richard A

; APPLICANT: Grosse, William M

APPLICANT: Szekeres, Edward S

; APPLICANT: Vernet, Corine A.M.

APPLICANT: LI, Li

APPLICANT: Stacie J. Casman, Stacie J. Boldo

APPLICANT: Boldog, Ferenc L  
APPLICANT: Gorman, Linda

APPLICANT: Gorman, Linda  
APPLICANT: Gancoll; Esba A

APPLICANT: Gangoili, Esha A  
APPLICANT: Fernandes, Elma B

APPLICANT: FERNANDES, ELMA K  
APPLICANT: RIEGER, DANIEL K

APPLICANT: Edinger, Shlomit R

APPLICANT: Gunther, Erik

APPLICANT: Millet, Isabelle

APPLICANT: Sciore, paul

APPLICANT: Ellerman, Karen

APPLICANT: MacDougall, John R

APPLICANT: Smithson, Glenda

; TITLE OF INVENTION: Proteins and

FILE REFERENCE: 21402-228

; CURRENT APPLICATION NUMBER: US/

; CURRENT FILING DATE: 2001-12-21

; PRIOR APPLICATION NUMBER: 60/25

; PRIOR FILING DATE: 2000-12-21

; PRIOR APPLICATION NUMBER: 60/258

PRIOR FILING DATE: 2000-12-20

;; PRIOR APPLICATION NUMBER: 60/286  
: PRIOR FILING DATE: 2001-03-20

;; PRIOR FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 60/274

;; PRIOR APPLICATION NUMBER: 60/271  
: PRIOR FILING DATE: 2001-03-08

; PRIOR FILING DATE: 2001-03-08

```
; PRIOR APPLICATION NUMBER: 60/277,826
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/279,840
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/282,981
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/283,656
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/309,247
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/311,754
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/313,331
; PRIOR FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 260
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 42
; LENGTH: 914
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-032-189-42

Query Match      17.8%; Score 851; DB 12; Length 914;
Best Local Similarity 28.0%; Pred. No. 1.2e-69;
Matches 256; Conservative 144; Mismatches 324; Indels 190; Gaps 33;

QY 38 YGYKVKVYKDSGTRSLTALPOLVKNSSVGGPDQILLSTASLESH-DRLRVRTDAXH 96
DB 43 YQALDSVTTDEDSTRF-----QIINEAS----KVPLLAEIYIGIEGNIFRLKINEETPLK 93
QY 97 RWEIPDNILHRHQPPPPPHSSLSLRYTLSSPTNR-----RKILLSHPNSDLTF 148
DB 94 PRFEVPD-----V-TSKPSTVRLSCSGDGLILADKGGLK 132

QY 149 SLINTTFF-----GFTSRKSTHVDLPF-----DATPDTNPN--- 180
DB 133 H-ITANPKFVDLSEEEVINSGLGYFEHLQ:LKQRAAKENBEETSVC*SQENQED 191
QY 181 -----TFLIFID-----QVLHITSLPGTRAHYGLGHSKPTQLAHNOTLTWRAAD 228
DB 192 LGLMEERKFGKVDIKANGPSSIGDLSLHGFPE-HLYGIPQHAEE-----SHQAKNTGGDA 245
QY 229 IPSSNPQV-----NLYGSHPFYMDVRSSPVAGSTHGVLLNSNGM-----DVE 271
DB 246 YRLYNLDVYGVQYDKMGYGVFPYL-----AHKLGRTIGIFMLNASETLVEINTEPAVE 301
QY 272 YT-----GNR--ITYKVIIGIIDLFFAGSPQGVVEQFTRVIGRPAZMPYV 316
DB 302 YTLTQMGPVAAKQKVGSRTHVHMSSEGIIDVFLLTGPTPSDVFKQYSHLTGTQAMPPULF 361
QY 317 AFGFOCRYGYHDVYELQSVVAGYAKAKIPLVNVMTDIDYMDAYKDFTLDPVNFPLDKKK 376
DB 362 SLGYHQCRWYEDQDVKAVDAGDEHDIPYDAWMLDIEHTGKRYFTWKNRFP--NPK 419
QY 377 KFNVLNHRKQKYVILDPGISTNKTETYIRGMKHDVFLK-RNGKPYLGSVNPQGVYFP 435
DB 420 RMQELLRSKRKLWISDPHIKIDPDYSVYVYKAKDQGFVKNQEGEDFEGVCWFGGLSSYL 479
QY 436 DFLKPSALTFTWDEIKRFLNLPV-DG-----LWIDMNEISNFISSPPPGSTLONPPY 468
DB 480 DPTPKVREWISS-----LFAPFYQGSTDILFLWDMNNEPSVF-RGP----- 521
QY 489 KINNSGVMLPIINKTIPPTAHYGDIDPEYVNVNLFGLVLEAVTRAAIKLUT--EKRPFFVL 546
DB 522 -----EQTMQKNAIHHGNWEHRELHNIYGFHQWATAEGLIKSKGKERPFVL 569
QY 547 SRSPFSGSKYTAHWGTGNAATWMDLVYSPMSLDGLFGIPMVGADICGFLGNTTEELC 606
DB 570 TRSFEPAGSQKYGAVTGTONTAWEWSNLKISIPMLLTLITGVSGFADIGGFIGNPETELL 629
QY 607 RRVICLGAFYPSRDHSLGTYQELRYW-RSVAAASAKVLGLRYTLPLPYFYLTMYEAQL 665
DB 630 VRWYQAGAYQFFFRGHATMNTKRRPEMLFGSEHTRLIREAIRERYGLPLWYSLFYHAHV 689
```

```
666 NGIPIARPLFPSPDDIKTYGISSQFLKGVVMSVPLKPGVVSVTAYFPRGN--WFDLF 723
690 ASQVMEPLWVEFPDELKTFDEMEYMLGSALLVHPVTEPKATVDVFLPGSNEVM--Y 746
724 DYTRSVTASTGRVYVTLASAPPDHINVHIOBNILAMO---GKAMTTOAARKTPPHLLVWMS 780
747 DYKTFAHMEGGCTVKIPVALDTIPVFORGSGSVIPKITTVGK--STGMMTESSYGLRVALS 804
781 DCGASFGEFLDDGVEVTMGVNGKMTFVKFIAASAKQTCIITSDVVUSGEFA-----VSQ 835
805 TKGSSVGEVLYDDGHSPOY-LHQKQFLHRKF-----SPCSSVLINSFADQRGHYFS 854
836 KWIWDKVTILGLRK 849
855 KCVVEKILVLGFRK 868

RESULT 8
US-10-032-189-40
; Sequence 40, Application US/10032189
; Publication No. US20030170630A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook II, John P
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Zernhusen, Bryan D
; APPLICANT: Patturajan, Meera
; APPLICANT: Grosse, William M
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Shimkets, Richard A
; APPLICANT: Grosse, William M
; APPLICANT: Szekeres, Edward S
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Li, Li
; APPLICANT: Casman, Stacie J
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Gorman, Linda
; APPLICANT: Gargolli, Esha A
; APPLICANT: Fernandes, Elma R
; APPLICANT: Rieger, Daniel K
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Gunther, Erik
; APPLICANT: Millet, Isabelle
; APPLICANT: Sciore, Paul
; APPLICANT: Ellerman, Karen
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glennda
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-228
; CURRENT APPLICATION NUMBER: US/10/032,189
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,495
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/258,171
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/269,940
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 60/274,192
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/277,826
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/279,840
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/282,981
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/283,656
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/309,247
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/311,754
```





```
Query Match      17.8%; Score 847; DB 12; Length 912;
Best Local Similarity 28.0%; Pred. No. 2.8e-69;
Matches 256; Conservative 142; Mismatches 324; Indels 192; Gaps 33;

QY 38 YGVKVKVSGSGTGRSLTALPOLVKNSSVGVQDIOQLISITASLESN-DRLRVRIITDAKH 96
Db 43 YRALDSVTTDEDSRFG-----QIINERAS-----KVPLLAEIYGIEGNIFRLKINEETPLK 93
QY 97 RWEIIPDNILHRHQPPPPPHSSLSLYTLSSPTTNR-----RKILLSHFNSDLTF 148
Db 94 PRFEVPD-----VLTSKSTVELISCSGDTGSLILADGKGLKC 132
QY 149 SLINTTFF-----GFTISRKSTHVDLF-----DAPDPTNPN--- 180
Db 133 H-ITANPFKVDLUSEEVVISNSGQLYFEHLQILHKORAAKENEETSVDTSQENQED 191
QY 181 -----TELIFID-----QYUHLTSSLPGRTRAHYVGLGHSKPTFOLAHNQTUTKAAAD 228
Db 192 LGLWEKFKCFDIKANGPSSIGLDFSLHGFE-HLYGIPQHAEE-----SHOLKNTGDAYR 245
QY 229 IPSSNDPV-----NLYGSHPTFMDVRSPVAGSTHGVLNLSNGM-----DVE 271
Db 246 L-YNLVDVGYQIDKMGYISVPYLL-----AHKLGRITIGIFWLNASETLVEINTEPAVE 299
QY 272 YT-----GNRIITYKIGIIDLFFAGSPGQVVEQFTRVIGRPAFMPYV 316
Db 300 YTLTQMPVAAKQKRSRTHVHMSSEGLIDVLLTGTPTSDVFKQYSHLTGTQAMPPLF 359
QY 317 AFGFOCRGYHDVYELOSVAVGAKAPIEYVMWTDIDYMDAYKDFLDVFNFFLDMKX 376
Db 360 SLGYHQCRWYDEQCVKAVDAGDFEHDIPYDAMWLDIEHTGKRYFTWQKNRFP--NPK 417
QY 377 KFWNLHKNQKQVYVLDIGISTNKTETETIRGMKHVDVFLK-RNGKPYLGSVWPGVYVP 435
Db 418 RQCELLRSKRLKLVISDPIHIEPDYSVYVAKDQGFVKNQEGEDFEGVCMPLSSYL 477
QY 436 DFLKPSALTFTWDEIKRFLNLLPV-DG-----LMDMNETSNFISSPPIPGSTLDPY 488
Db 478 DETNPKRVREWYSS-----LFAPPVYQGSSTDLFLWMDMNEPSVF-RGP----- 519
QY 489 KINNSGWLPIINKTIPPTAMHYGDIPEYVNHNLFGYLEARVTRAAATIKLT--EKRPFVL 546
Db 520 -----EQTMQKNALHGHGWEHRECHNIYGFYHQMATAGLLKRSKGRERFVL 567
QY 547 SRSTFGSKYTAHTWGDNAATWDLVYSIPSMDFGLFGIPMVGADICGFLGNTTEELC 636
Db 568 TRSFFAGSKYCAVWVGONTAEWSNLKISIPMLLTSLTISGIFSGADIGGFIQNPETELL 627
QY 607 RKNVQLGAFYFESRSHSLGTTTYOELYRK-ESVAASARKVLGLRYTLPLPYFTLMEYQAL 665
Db 628 VRWYQAGATQPFPRGHATWTKRRPFLFGESHTLREARIRYGLLPYWSLFYRAHY 687
QY 666 NGIPTARPLFFSPDDIKTYGISSOFLKGGVWSPVLKPGVWVTAFFPRGN--WFDLF 723
Db 688 ASOPVMRPLWVEFPDELKTFDMEDEYMLGSAALLVHPVTEPKATTVDVFLPGSNEVW--Y 744
QY 724 DYTRSVTASTGYVYLSAPPDINHVIQECNLANQ-----GKAMTQAAKTPFELLVMS 780
Db 745 DYKTFAHWEGGCTVPIPVALDITIPVQGGGVPIKTTVGK--STGWMTESSYGLRVALS 802
QY 781 DCGASFGELFDGVDVMTGVNRKWTFFVKFTAASAKQTCIITSDVVSGEFA-----VSQ 835
Db 803 TQGSVSGELYLDGHSFQY-LHQKQFLHRKP-----SFCSSVLNSPADQRGHYPS 852
QY 836 KWDVKTILGLRK 849
Db 853 KCWVEXILVGLFRK 866

RESULT 10
US-10-032-189-123
; Sequence 123; Application US/10032189
; Publication No. US20030170630A1
```

```
GENERAL INFORMATION:
; APPLICANT: Alsobrook II, John P
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Zernusen, Bryan D
; APPLICANT: Patturajan, Meera
; APPLICANT: Grosse, William M
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Shimkets, Richard A
; APPLICANT: Grosse, William M
; APPLICANT: Szekeres, Edward S
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Li, Li
; APPLICANT: Casman, Stacie J
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Gorman, Linda
; APPLICANT: Gangolli, Beha A
; APPLICANT: Fernandes, Elma R
; APPLICANT: Rieger, Daniel K
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Gunther, Erik
; APPLICANT: Millet, Isabelle
; APPLICANT: Sciore, Paul
; APPLICANT: Ellerman, Karen
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glennda
; FILE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-228
; CURRENT APPLICATION NUMBER: US/10/032,189
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,495
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/258,171
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/269,940
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 60/274,192
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/277,826
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/279,840
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/282,981
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/283,656
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/309,247
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/311,754
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/313,331
; NUMBER OF SEQ ID NOS: 260
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 123
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-032-189-123

Query Match      16.9%; Score 805.5; DB 12; Length 943;
Best Local Similarity 29.9%; Pred. No. 2.1e-65;
Matches 230; Conservative 108; Mismatches 284; Indels 147; Gaps 27;

QY 192 LTSSLPGTRAHYVGLGHSKPTFOLAHNQTUTRAADIPSS-----ADNRLKVTGEGEPYRLNLDVQYELYNP-MALY 239
Db 245 LDFSLPGME-HVYGIPEH-----ADNRLKVTGEGEPYRLNLDVQYELYNP-MALY 295
QY 240 GSHFPYMDVRSSPVAGSTH---GVLLNSNGMDVEYTGNRITYKVG----- 283
Db 236 GSVFVLL-----AHNPHRDLGIFWLNAAETWDISNTAGTTLFGKMMDYLQSGGETP 348
```

Qy 284 -----GIIDLYFFAGSPGVVEQFTRVIGRPAPMPYWAFFGQCRGYGHDVVEL 333  
Db 349 QTOVRWMSSETGIIDVFLLLGPSISDFRQYASLTGTQALPPLFSLGHOSRWMYRDEADV 408  
Qy 334 QSVVAGYAKAKIPLVNMWTDIDYMDAYKDFTLDPVNFPLDKMKKFVNNLHGKQKVVVIL 393  
Db 409 LEVDQGFDDHNLPCDVIWLDIEHADGKRYFTWDPSPRFP--QPRTMLERLASKERKLVAV 466  
Qy 394 DPGISTNKTYETVIRGMKHVDVFLK-RNGKPYLGSVMPGPVYFDFLKPALSALTFTWDEIKR 452  
Db 467 DPHIKVDSGTRVHEELNGLYKTRDGSYEGWCWPGSAGYDFDFTNPTWRAWMA----- 521  
Qy 453 FLNLLPVDG-----LWIDMNEISNFISSPPIPGSTLDNPPYKINNSGVMLPIINTK 503  
Db 522 --NMFSDYNEGSAPNLFVNDMNEPSVF-NGPEV-----T 554  
Qy 504 IPTAMHYGDIPEYVNNLFGYLEARVTRAAALIKLT--EKRPFLVSRSTFSSGKYTAHM 561  
Db 555 MLKDAQHYGWEHRDVHNIYGLYVHMATADGLRQSGGMRPFLARAFAGSQRFGAVM 614  
Qy 562 TGDNAATMNDLVYSIPMSLDFGLGIPMVGADICGELGNTTELCRWWICLGFYPPFSRD 621  
Db 615 TGDNTAEWDLKISIPNCLSLGLVGLSFCGADVGFGFKNPEPELLVRYQMGAYQFFFA 674  
Qy 622 HSSLGTYQELYRWESVAAS-ARKVLGLRYTLPLPYFTLMYEAOQNGIPIARPLFFSFPD 680  
Db 675 HAHLDTCRRSPWLLPSQHNDIIRDALGQRYSLIPFWYTLLYQAHREGIPWRLPLWQYQ 734  
Qy 681 DIKTYGISSOFELCKGWNVPVLKGVVSVTATFP-RGN-WFLLFDYTRSVTASTGRY-- 736  
Db 735 DVTFNIDDOYLLGDALLHPVSDSGAHGQVQVLPQCGGEVWYDIQSYQKHGPQT-LYLP 793  
Qy 737 VTLSPAPDHNVHIOEGNILLAQCKA-MTTOAARKTPHLLVVMVSDGASFGELFLDDGV 795  
Db 794 VTLSSIP-----VFORGTIVPRWVRVSSECKMDDPITLFLVALSPQIAQELFLDDG- 848  
Qy 796 EVTMVNRGKWTVPKFIASAKQTCIITSQVSGEFAVSQK-----WVIDKVTI 844  
Db 849 -----VTF-----NYOTROEFLRLRRFSFGNTLVSSADPEGHFETPIW-IERVVI 893  
Qy 845 LGLRKGTKINGYVTRTCAVTRKGD-KSKLASTDRKGEFIVABISGLNL 892  
Db 894 IGACKPAV-----VLQTKGSPESRLSFQHDPETSVLVRKPGINV 934

## RESULT 11

US-10-032-189-122

Sequence 122, Application US/10032189

Publication No. US20030170630A1

## GENERAL INFORMATION:

APPLICANT: Alsobrook II, John P  
APPLICANT: Tchernev, Velizar T  
APPLICANT: Liu, Xiaohong  
APPLICANT: Spytek, Kimberly A  
APPLICANT: Zerhusen, Bryan D  
APPLICANT: Patturajan, Meera  
APPLICANT: Grosse, William M  
APPLICANT: Lepley, Denise M  
APPLICANT: Burgess, Catherine E  
APPLICANT: Shimkets, Richard A  
APPLICANT: Grosse, William M  
APPLICANT: Szekeres, Edward S  
APPLICANT: Vernet, Corine A.M.  
APPLICANT: Li, Li  
APPLICANT: Casman, Stacie J  
APPLICANT: Boldog, Ferenc L  
APPLICANT: Gorman, Linda  
APPLICANT: Gangolli, Esha A  
APPLICANT: Fernandes, Elma R  
APPLICANT: Rieger, Daniel K  
APPLICANT: Edinger, Shlomit R  
APPLICANT: Gunther, Erik

APPLICANT: Millet, Isabelle  
APPLICANT: Sciore, Paul  
APPLICANT: Ellerman, Karen  
APPLICANT: MacDougall, John R  
APPLICANT: Smithson, Glennda  
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-228  
CURRENT APPLICATION NUMBER: US/10/032,189  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: 60/257,495  
PRIOR FILING DATE: 2000-12-21  
PRIOR APPLICATION NUMBER: 60/258,171  
PRIOR FILING DATE: 2000-12-20  
PRIOR APPLICATION NUMBER: 60/269,940  
PRIOR FILING DATE: 2001-02-20  
PRIOR APPLICATION NUMBER: 60/274,192  
PRIOR FILING DATE: 2001-03-08  
PRIOR APPLICATION NUMBER: 60/277,826  
PRIOR FILING DATE: 2001-03-22  
PRIOR APPLICATION NUMBER: 60/279,840  
PRIOR FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: 60/282,981  
PRIOR FILING DATE: 2001-04-11  
PRIOR APPLICATION NUMBER: 60/283,656  
PRIOR FILING DATE: 2001-04-13  
PRIOR APPLICATION NUMBER: 60/309,247  
PRIOR FILING DATE: 2001-07-31  
PRIOR APPLICATION NUMBER: 60/311,754  
PRIOR FILING DATE: 2001-08-17  
PRIOR APPLICATION NUMBER: 60/313,331  
PRIOR FILING DATE: 2001-08-17  
NUMBER OF SEQ ID NOS: 260  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 122  
LENGTH: 944  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-032-189-122

Query Match 16.98; Score 805.5; DB 12; Length 944;

Best Local Similarity 29.98; Pred. No. 2.2e-85;

Matches 230; Conservative 108; Mismatches 284; Indels 147; Gaps 27;

Qy 192 LTSSLPGTRAHYGLGHSKPTFLAHNQTLTWRAADIPSS-----NPDVNLV 239  
Db 246 LDFSLPGME-HVYGIPEH-----ADNLRKLVTEGEPVRLYNLDVFOVELYNP-MALV 296  
Qy 240 GSHPPYMDVRSSPVAGSTH---GVLLNSNGMDVEYTGNRITYKVG----- 283  
Db 297 GSVPEVLL-----AHNPHRDLGIFWLNAAETWVDISSNTAGTKLFGKMDYLQSGGETP 349  
Qy 284 -----GIIDLYFFAGSPGVVEQFTRVIGRPAPMPYWAFFGQCRGYGHDVVEL 333  
Db 350 QTOVRWMSSETGIIDVFLLLGPSISDFRQYASLTGTQALPPLFSLGHOSRWMYRDEADV 409  
Qy 334 QSVVAGYAKAKIPLVNMWTDIDYMDAYKDFTLDPVNFPLDKMKKFVNNLHGKQKVVVIL 393  
Db 410 LEVDQGFDDHNLPCDVIWLDIEHADGKRYFTWDPSPRFP--QPRTMLERLASKERKLVAV 467  
Qy 394 DPGISTNKTYETVIRGMKHVDVFLK-RNGKPYLGSVMPGPVYFDFLKPALSALTFTWDEIKR 452  
Db 468 DPHIKVDSGTRVHEELNGLYKTRDGSYEGWCWPGSAGYDFDFTNPTWRAWMA----- 522  
Qy 453 FLNLLPVDG-----LWIDMNEISNFISSPPIPGSTLDNPPYKINNSGVMLPIINTK 503  
Db 523 --NMFSDYNEGSAPNLFVNDMNEPSVF-NGPEV-----T 555  
Qy 504 IPTAMHYGDIPEYVNNLFGYLEARVTRAAALIKLT--EKRPFLVSRSTFSSGKYTAHM 561  
Db 556 MLKDAQHYGWEHRDVHNIYGLYVHMATADGLRQSGGMRPFLARAFAGSQRFGAVM 615  
Qy 562 TGDNAATMNDLVYSIPMSLDFGLGIPMVGADICGELGNTTELCRWWICLGFYPPFSRD 621

616 TGDNTAENDHLKISIPMCISLGLVLSFCGADVGCFKPNPELVVRWQMGAYQPFRA 675  
 622 HSSLTGYOELRVESVNAS-ARKVLGLRYTLLPYFTLMEYEAQNGIPIARPLFFSPFD 680  
 676 HAHLDTCRRPEWLLPSQHNDIRDALGORYSLLPEFTLLYQAHREGIPVWRPLWVQYPQ 735  
 681 DIKTYGSSQFLKGVVWSPVLKPGVVSUYATFP-RGN-WFDLFDVTRSVTASTGRY-- 736  
 736 DVTTFNDQVLLGDALLVHPVSUSGAGVGVVLPQGEVWYDIQSKHGHQPT-LYLP 794  
 737 VTLAPPDHHVHQBNILAMQGA-MTTOAARKTFFHLLVWMSDCGASFGELFDDGV 795  
 795 VTLSSIP---VFQGGTIVPRWVRSSSECMKDDPITLFFVALSPQTAQGFLLDDG- 849  
 796 EVTWGVNKGKTFVKFPIAASAKQTCIITSVVVSSEFAVSQK-----WVIDKVTI 844  
 850 -----YTF-----NYQTRQEFLLRRFSFGNTLVSSADPEGHFETPIW-IERV 894  
 845 LGLRKGTKINGYVRTGAVTRKGD-KSKLKSTPDRKGEFIVAEISGLNL 892  
 895 IGACKPAV-----VLQTKGSPESLSFQHDPEITSVLVLRKPGINV 935  
 RESULT 12  
 US-10-032-189-121  
 ; Sequence 121, Application US/10032189  
 ; Publication No. US20030170630A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Alsbrook II, John P  
 ; APPLICANT: Tchernev, Velizar T  
 ; APPLICANT: Liu, Xiaohong  
 ; APPLICANT: Spytek, Kimberly A  
 ; APPLICANT: Zethusen, Bryan D  
 ; APPLICANT: Patturajan, Meera  
 ; APPLICANT: Grosse, William M  
 ; APPLICANT: Lepley, Denise M  
 ; APPLICANT: Burgess, Catherine E  
 ; APPLICANT: Shimkets, Richard A  
 ; APPLICANT: Grosse, William M  
 ; APPLICANT: Szekeres, Edward S  
 ; APPLICANT: Vernet, Corine A.M.  
 ; APPLICANT: Li, Li  
 ; APPLICANT: Casman, Stacie J  
 ; APPLICANT: Boldog, Ferenc L  
 ; APPLICANT: Gorman, Linda  
 ; APPLICANT: Gangolli, Esha A  
 ; APPLICANT: Fernandes, Elma R  
 ; APPLICANT: Rieger, Daniel K  
 ; APPLICANT: Edinger, Shlomit R  
 ; APPLICANT: Gunther, Erik  
 ; APPLICANT: Millet, Isabelle  
 ; APPLICANT: Sciore, Paul  
 ; APPLICANT: Ellerman, Karen  
 ; APPLICANT: MacDougall, John R  
 ; APPLICANT: Smithson, Glenda  
 ; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
 ; FILE REFERENCE: 21402-228  
 ; CURRENT APPLICATION NUMBER: US/10/032,189  
 ; CURRENT FILING DATE: 2001-12-21  
 ; PRIOR APPLICATION NUMBER: 60/257,495  
 ; PRIOR FILING DATE: 2000-12-21  
 ; PRIOR APPLICATION NUMBER: 60/258,171  
 ; PRIOR FILING DATE: 2000-12-20  
 ; PRIOR APPLICATION NUMBER: 60/269,940  
 ; PRIOR FILING DATE: 2001-02-20  
 ; PRIOR APPLICATION NUMBER: 60/274,192  
 ; PRIOR FILING DATE: 2001-03-08  
 ; PRIOR APPLICATION NUMBER: 60/277,826  
 ; PRIOR FILING DATE: 2001-03-22  
 ; PRIOR APPLICATION NUMBER: 60/279,840  
 ; PRIOR FILING DATE: 2001-03-29  
 ; PRIOR APPLICATION NUMBER: 60/282,981  
 ; PRIOR FILING DATE: 2001-04-11

; PRIOR APPLICATION NUMBER: 60/283,656  
 ; PRIOR FILING DATE: 2001-04-13  
 ; PRIOR APPLICATION NUMBER: 60/309,247  
 ; PRIOR FILING DATE: 2001-07-31  
 ; PRIOR APPLICATION NUMBER: 60/311,754  
 ; PRIOR FILING DATE: 2001-08-17  
 ; PRIOR APPLICATION NUMBER: 60/313,331  
 ; PRIOR FILING DATE: 2001-08-17  
 ; NUMBER OF SEQ ID NOS: 260  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 121  
 ; LENGTH: 966  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 ; US-10-032-189-121  
 Query Match 16.9%; Score 805; DB 12; Length 966;  
 Best Local Similarity 28.1%; Pred. No. 2.5e-65;  
 Matches 275; Conservative 133; Mismatches 369; Indels 200; Gaps 41;  
 QY 12 ILLVFLQYL---VAGISTSENDPEGVIGYGVKVKV-KV-DSGTTR----- 53  
 DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 85 VLLVLEQLQKQNMTRIRIDELEPRP---RYRVPDVLVADPPTARLSVSGRDONSVELT 141  
 QY 54 -----SLTALP---QLVKNSSVYGPDIOLLSITA-----SLESNDLRYITD---A 94  
 DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 142 VAEGYKILTAQPPRLDLEDRSL-----LLSVNARGMLMAFEHQRAPRVPFSDKVS 195  
 QY 95 KHRWEIPDNILHRH-----QPPPPPHSLSLYRTLSSPTTNRKILLSPN 143  
 DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 196 LGSVMDKIKNLFQRQESKDPAGNGAQFEATPGDG-----DKPEETQEKAEKDEFG 246  
 QY 144 S-DLTFSLINTPFGFTISRKSTHDVLPDATP-DPTNPNTFLIFIDQVHLTSSLPQTRA 201  
 DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 247 AWEETP-----KTHS---DSKPYGTS-----VGLDFSLPGMB- 276  
 QY 202 HIYGLGEHS-----KPTFLQAHNOTLTNRADIPSSNPDVNLVGS-----HPFVMDVR- 249  
 DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 277 HVYGIPEHADSRLKVTGEGEYRLYNLDVQYELNRP-MALYGSVPVLLAHSFRLDGI 335  
 QY 250 -----SSPVAGST-HGVLL--LNSNGMDVEYTNRNITYKVIGGIDLYFFAGP 294  
 DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 336 FWLNAETWVDISSNTAGTKTLFGKMLDYLGSG---ETPQTDIRWMSSEGIIDVFLMLGP 392  
 QY 295 SPQGVVEQFTRVIGEPAPMPYNAFGQCRGYHDVLOSVAQYAKAKIPLVEMWTDI 354  
 DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 393 SVDFVFRYASLTGTQALPPLFSLSGHOSRWNYRDEADVLEVDQGFDDHNPCCDVIWLDI 452  
 QY 355 DYMDAYKDFTLDPVNFPLDKMKKFVNNLHKQKQVYVILDPGISTNKTYETVIRGMKH 414  
 DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 453 EHADGKRYFTWDPTRFP--QPLNMLEHLASKRKLVAIVDPHIKVDGSGYRVHEELRNHGL 510  
 QY 415 FLK-RNGKPYLGSVWPGVPYFPDFLKPSALTFTWDEIKRFLNLLPVDG-----LWI 464  
 DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 511 YVKTRDGSYEGWCPGASYPDFTNPRMAWS-----NMFSEFNYEGSAPNLVYN 563  
 QY 465 DMNEISNFISSPPIPGSLTLDNPPYKINNSGVMLPIINKTIPPTAMGHYGDIPYVNHVLF 524  
 DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 564 DMNEFSVF-NGPEV-----TMLKDAVHYGGWHRDHIY 598  
 QY 525 YLEARVTRAALIKLTE--KRPVLSRSTFGSGKYTAHTWGTGDNAAATWDLVYSPMLDF 582  
 DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 599 LYVHMATADGLIQRSGGIERFVLBSRAFFSGQRFAGVMTGDNATDNDHLSIPMCLSL 658  
 QY 583 GLFGIPMVGADICGFLGNTTSELCCRWTQLGAFYFSSRDHSSLSGTTTQOELRWESVAASA 642  
 DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 659 ALVGLSFCGADVGGFFKNPEPELLVRWYQMGAYQPFRAHAHLDTGRREPWLLASQYQDA 718  
 QY 643 -RKVLGLRYTLLPYRTLMYEAQLNGIPIARPLFSPDDIKTYGISSQFLGKGVWVSP 701  
 DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 719 IKDALFQRIYSLLPFWYTLFQAHKEGFPVMPPLWQVYPEDMSTFSIEDQFLLDALLHP 778  
 QY 702 VLKPGWSVTAYFPRGN--WFDLFDYTRSVTASTGRY--VTLAPPDHHVHQBNILA 757



QY 839 IDKVTILGLRKGKINGYTVTGAVTRKGD-KSKLAKSTPORKGEFIVAEISGLNL 892  
Db 911 IERVIIIGAGKPAV-----VLQTKGSPESRLSFQHPFETSVLVRKPGINV 957

RESULT 14

US-10-032-189-124  
; Sequence 124, Application US/10032189  
; Publication No. US20030170630A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsobrook II, John P  
; APPLICANT: Tchernev, Velizar T  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Zernhusen, Bryan D  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Grosse, William M  
; APPLICANT: Lepley, Denise M  
; APPLICANT: Buigess, Catherine E  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Grosse, William M  
; APPLICANT: Szekeres, Edward S  
; APPLICANT: Vernet, Corine A.M.  
; APPLICANT: Li, Li  
; APPLICANT: Casman, Stacie J  
; APPLICANT: Boldog, Ferenc L  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gangolli, Esha A  
; APPLICANT: Fernandes, Elma R  
; APPLICANT: Rieger, Daniel K  
; APPLICANT: Edinger, Shlomit R  
; APPLICANT: Gunther, Erik  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Sciore, Paul  
; APPLICANT: Ellerman, Karen  
; APPLICANT: MacDougall, John R  
; APPLICANT: Smithson, Glenda  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-228  
; CURRENT APPLICATION NUMBER: US/10/032,189  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 60/257,495  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/258,171  
; PRIOR FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: 60/269,940  
; PRIOR FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 60/274,192  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/277,826  
; PRIOR FILING DATE: 2001-03-22  
; PRIOR APPLICATION NUMBER: 60/279,840  
; PRIOR FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/282,981  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: 60/283,656  
; PRIOR FILING DATE: 2001-04-13  
; PRIOR APPLICATION NUMBER: 60/309,247  
; PRIOR FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: 60/311,754  
; PRIOR FILING DATE: 2001-08-17  
; PRIOR APPLICATION NUMBER: 60/313,331  
; PRIOR FILING DATE: 2001-08-17  
; NUMBER OF SEQ ID NOS: 260  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 124  
; LENGTH: 944  
; TYPE: PRT  
; ORGANISM: Sus scrofa  
US-10-032-189-124

Query Match 16.8%; Score 799; DB 12; Length 944;  
Best Local Similarity 26.1%; Pred. No. 8.7e-65;

Matches 264; Conservative 148; Mismatches 374; Indels 224; Gaps 39;  
QY 10 LGILLVFLLOVLVAGISTSENDPEGVIGYGYKVKSVKVDSTGTRRSALTALPOLVKNSVYG 69  
Db 23 LGVCLGLTLAVDRSNFKTCES-----SPCKQRQSIRPGQSPYRALLDSLQL-----G 70  
QY 70 PD-----IQLLSITASLESNDRLVRITDAXHRRWEIPONILHRHOPPPPP 116  
Db 71 PDTLTIHLINVTKVLVLELQGLQKNMTRIRIDELEPRPRYRVPDLV---AEPETAR 127  
QY 117 HSLS-----SLVRTLLSPTTNRKRILLSHSPNSDLTFSLINTTTPFGFTISRKSTHDVLFD 171  
Db 128 LSVSGODNSVEVTAEGP---YKILLTARP---FRDLLLEDRSLLSVNARGLLNFHQ 181  
QY 172 ATP-----DPTN-----PNTFLIFIDO----- 188  
Db 182 RAPRVSGSKDPAEGDGAQPEEAPGDGDKPBETIOGKARKDEPGAWEETFKTHSDSKPYG 241  
QY 189 -YLHLTSSILPGTRAHIYGLGHSKPTFFOLAHNQTLTWRAAD-----IFSSN 233  
Db 242 TSVGLDFFSLPGME-HVIGIPEHA-----DSURLKTEGGDPYRLYNLDVFOYELYN 291  
QY 234 PDVNLVYGSHPFYMDVRSSPVAGSTH---GVLLLSNCGMDVEYTGNRITYKIVG----- 283  
Db 292 P-MALYGSVPVLL-----AHSPHRDLGIPWLNAAEETWVDISSNTAGKTTLFGKMLDYLQ 343  
QY 284 -----GIIDLYFFAGSPSGQVQFTRVIGRPAPMPYWAQFQOCRYG 327  
Db 344 GSGETPQTDVRMMSEGIIDVFLGLGPSVDFVFRQYASLTGTQALPPLFSLYGHQSRWNY 403  
QY 328 HDVVELQSVVAGYAKAKIPLEVMWTDIDYMDAYKDFTLDPVNFPLDKMKKPVNLLKNGQ 387  
Db 404 RDEADVLEVNQFDDHNLPCDFIWLIDIEHADGKRYFTWDPSEFP--QPRTMLEHLASRR 461  
QY 388 KYVVLDPGISTNKTYETIRGMKHDVFLK-RNGKPYLGSVMPGPGVYFPDFLKFSAITFW 446  
Db 462 KLVAIVDPHIKVDSSYRVHEELQNLGLVYKTRDGSDEGMCWPGAASYPDFTNPKMRAW 521  
QY 447 TDEIKRFLNLLPVDG-----LWIDMNEISFISSPPIPGSTLONPPYKINNNGVMLPII 500  
Db 522 AD-MFRFEN---YEGSSNLVYVNDMNEPSVF-NGPEV----- 554  
QY 501 NKTIPPTAMHYGDIPEYVNVHNLFGYLEARVTRAALIKLT--EKRPVLSRSTFSGSKYT 558  
Db 555 --TMLKDAQOHYGGWEHRLHNIYGFYVHMATADGLVLSGGVERPFVLSRAFFAGSORFG 612  
QY 559 AHWTDGNAATNDLVYSTPSMLDPGLFGIPMVGADICGFLGNTTEELCRRWQLGAFYFP 618  
Db 613 AVMTGDNTAEMDHLKISIPMCLSLGLVGVSGADVGFFKNPEPELLVRYOMGAYQPF 672  
QY 619 SRDHSLSG-----TTYQELYRMESVAASARKVLGLRYTLPLYFTILMYEQLNGI 668  
Db 673 FRAHAHLDTGRREPWLLPTQYQDM-----IRDALGQYSILLFPWITLIFYQHREGV 723  
QY 669 PIARPLFFSPDDIKTYGISSQFLGKGMVSPVLKPGVSWVTAYFP-RGN-WFDLFDYT 726  
Db 724 PVMRALWVHYPODVTTFSDIDDFLLGDALLVHPVTDSEAHGVQVYVLPQGGEVWYDVHSYQ 783  
QY 727 RSVTASTGRY---VTLSAPDPDHNVHIQEGNILAMOGKA-MTQAARKTPFHLVNVMSDCG 783  
Db 784 KYHGFQT-LYLFVTLSSIP---VFQGGTIVPRMVRSSDCMKDDPITLFLVALSPQG 838  
QY 784 ASFGELFLDDGVEVTGMVNRGKMTFVKFAASAKQTCIITSVVSGEFAVSKQWIDKVT 843  
Db 839 TAQGEFLFDDG--HTFNYQTGHEFLRRFSFG-NLVSSADSCKGHE-TPVM-IERVV 893  
QY 844 ILGLRKGKINGYTVTGAVTRKGD-KSKLAKSTPORKGEFIVAEISGLNL 892  
Db 894 IIGAGKPA-----TVVLQTKGSPESRLSFQHPDTPSVLILRKPGVNV 935

RESULT 15  
US-10-032-189-38

```
Sequence 38, Application US/10032189
Publication No. US20030170630A1
GENERAL INFORMATION:
APPLICANT: Alsbrook II, John P
APPLICANT: Tchernev, Velizar T
APPLICANT: Liu, Xiaohong
APPLICANT: Spytek, Kimberly A
APPLICANT: Zerhusen, Bryan D
APPLICANT: Patturajan, Meera
APPLICANT: Grosse, William M
APPLICANT: Lepley, Denise M
APPLICANT: Burgess, Catherine E
APPLICANT: Shimkets, Richard A
APPLICANT: Grosse, William M
APPLICANT: Szekeres, Edward S
APPLICANT: Vernet, Corine A.M.
APPLICANT: Li, Li
APPLICANT: Casman, Stacie J
APPLICANT: Boldog, Ferenc L
APPLICANT: Gorman, Linda
APPLICANT: Gangolli, Esha A
APPLICANT: Fernandes, Elma R
APPLICANT: Rieger, Daniel K
APPLICANT: Edinger, Shomit R
APPLICANT: Gunther, Erik
APPLICANT: Millet, Isabelle
APPLICANT: Sciore, Paul
APPLICANT: Ellerman, Karen
APPLICANT: MacDougall, John R
APPLICANT: Smithson, Glenda
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-228
CURRENT APPLICATION NUMBER: US/10/032.189
CURRENT FILING DATE: 2001-12-21
PRIORITY FILING DATE: 2000-12-21
PRIORITY FILING DATE: 2000-12-21
PRIORITY FILING DATE: 2000-12-20
PRIORITY FILING DATE: 2000-12-20
PRIORITY FILING DATE: 2001-02-20
PRIORITY FILING DATE: 2001-02-20
PRIORITY FILING DATE: 2001-03-08
PRIORITY FILING DATE: 2001-03-22
PRIORITY FILING DATE: 2001-03-29
PRIORITY FILING DATE: 2001-04-11
PRIORITY FILING DATE: 2001-04-13
PRIORITY FILING DATE: 2001-07-31
PRIORITY FILING DATE: 2001-08-17
PRIORITY FILING DATE: 2001-08-17
PRIORITY FILING DATE: 2001-08-17
NUMEROUS OF SEQ ID NOS: 260
SOFTWARE: Patent in Ver. 2.1
SEQ ID NO 38
LENGTH: 967
TYPE: PRT
ORGANISM: Homo sapiens
US-10-032-189-38

Query Match 16.7% Score 797.5; DB 12; Length 967;
Best Local Similarity 25.6%; Pred. No. 1.2e-64;
Matches 266; Conservative 156; Mismatches 365; Indels 251; Gaps 36;

QY 19 QYLVAGISTSENDE-GVIGYGYKYSK------VDSGTRRSALTALPOLVKN 65
Db 26 QALLDSVTDEDSTRFOINEASKSRQKWLKSKSTYQALLDSVTDEDSTRFO:INEA 85
QY 66 SVYGPDIQLLS:TSLESN-DRLURVITDAKURRWEIPDNILHRHQPDPHPHS-SSLYR 124
```

Search completed: October 27, 2003, 10:38:15  
Job time : 112.481 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 ; Search time 16.3223 Seconds  
(without alignments)  
2340.767 Million cell updates/sec

Title: US-10-043-418-3  
Perfect score: 4768  
Sequence: 1 MKKKIPSLALGILVFLQY.....VAEISGLNLLGRFPLVILH 903

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/prodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/prodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/prodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/prodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2422.5	50.8	877	1	US-08-430-925A-4
2	831	17.4	914	4	US-09-437-054A-8
3	811.5	17.0	919	4	US-09-437-054A-17
4	676	14.2	693	4	US-09-376-343-2
5	643.5	13.5	938	4	US-08-897-843A-1
6	487	10.2	1070	2	US-08-633-770A-2
7	476	10.0	1066	2	US-08-633-770A-1
8	371.5	7.8	1092	3	US-09-375-608-3
9	352.5	7.4	1088	3	US-08-633-768A-1
10	351	7.4	1091	3	US-08-633-768A-2
11	335	7.0	570	4	US-09-275-608-4
12	199	4.2	390	4	US-09-634-238-331
13	190.5	4.0	150	4	US-09-437-054A-10
14	130.5	2.7	2446	2	US-08-551-356-2
15	130.5	2.7	2446	2	PCT-US93-12687-2
16	126.5	2.7	2327	6	5455158-1
17	124.5	2.6	2386	2	US-09-016-366A-12
18	120.5	2.5	2324	1	US-08-283-857-1
19	120.5	2.5	2324	5	PCT-US95-09819-1
20	118.5	2.5	2231	1	US-08-153-799-16
21	114	2.4	286	4	US-09-634-238-330
22	113.5	2.4	1268	4	US-08-506-296B-28
23	109.5	2.3	1058	4	US-09-252-991A-29105
24	109.5	2.3	1170	1	US-08-032-364-2
25	108.5	2.3	632	4	US-09-853-533A-8
26	105.5	2.2	963	1	US-08-537-002A-3
27	105.5	2.2	963	3	US-08-863-010-3

28	105.5	2.2	963	3	US-09-024-429-3	Sequence 3, Appli
29	105	2.2	307	4	US-08-506-296B-63	Sequence 63, Appl
30	105	2.2	545	4	US-08-506-296B-75	Sequence 75, Appl
31	104.5	2.2	628	4	US-08-107-532A-5288	Sequence 5288, Ap
32	104	2.2	1025	2	US-08-304-309-2	Sequence 2, Appli
33	104	2.2	1025	3	US-08-991-942-2	Sequence 2, Appli
34	104	2.2	1025	3	US-09-138-103-2	Sequence 2, Appli
35	104	2.2	1025	5	PCT-US95-04567-4	Sequence 4, Appli
36	104	2.2	1231	4	US-09-071-035-420	Sequence 420, App
37	104	2.2	1265	4	US-09-071-035-418	Sequence 418, App
38	102	2.1	686	3	US-03-306-922-2	Sequence 2, Appli
39	102	2.1	1042	3	US-08-928-361B-11	Sequence 11, Appl
40	102	2.1	1042	4	US-09-588-995A-11	Sequence 11, Appl
41	102	2.1	1837	3	US-08-928-361B-5	Sequence 5, Appli
42	102	2.1	1837	4	US-09-588-995A-5	Sequence 5, Appli
43	101	2.1	574	3	US-09-383-586-36	Sequence 36, Appl
44	101	2.1	917	4	US-08-259-451-11	Sequence 11, Appl
45	100	2.1	531	3	US-08-688-988-39	Sequence 39, Appl

ALIGNMENTS

RESULT 1  
US-08-430-925A-4  
; Sequence 4, Application US/08430925A  
; Patent No. 5763252  
; GENERAL INFORMATION:  
; APPLICANT: Skadsen, Ronald W  
; APPLICANT: Tibbot, Brian K  
; TITLE OF INVENTION: Cloned Alpha-Glucosidase from Barley  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: 1 South Pinckney Street  
; CITY: Madison  
; STATE: WI  
; COUNTRY: US  
; ZIP: 53703  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/430,925A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J  
; REGISTRATION NUMBER: 27,386  
; REFERENCE/DOCKET NUMBER: 960296.93171  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 608-251-5000  
; TELEFAX: 608-251-9166  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 877 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-430-925A-4

Query Match 50.8%; Score 2422.5; DB 1; Length 877;  
Best Local Similarity 53.6%; Pred. No. 2.6e-218;  
Matches 486; Conservative 132; Mismatches 234; Indels 55; Gaps 15;  
QY 10 LGILVFLQLYLVAQ--ISTENDP---EGVIGYGYKYK-SVKVDSGT-RRSLTALPOLV 62  
Db 4 VGVLLLCCLCLFAPRLCSKEEGFLAARTVLAVAVTMEGALRAEATGGRSSTG----- 58  
QY 63 KNSVVGPDIOQLLSATLESNDRLRVITDAKRRWIPDNILRHOP-----PPPPPHS 118

Db 59 -----DVORLAVASLETDSRLVRITDADHPRMEVQDIIIPRPAQGVLDHAPPAS 110  
Qy 119 LSSLYRTLSSPTNRRKILLSHNSDLTSLINTTPEGFTISRKSTHDLFDATDPTN 178  
Db 111 -----SAPLOGR---VUSPAGSLVLT-VHASPFRTVSRSTGDTLFDAPG--- 154  
Qy 179 PNTFLIFIDQYLHLTSSLPGTRAH:YGLGEHSKPTFQLAHNQTULTMRAADIPSNPDVNL 238  
Db 155 ----LVFRDKYLEVTSALPAGRASLYLGEHTKSSFLRHNDSTFLWNAIGASVYCVNL 210  
Qy 239 YGSHPPFMDVRSSVAGSTHGVLLNNGMDVEYTGNRITYKVIIGIIDLYFFRAGSPGQ 298  
Db 211 YGSHPPFMDVRA---PGTAHGVLSSNGMDLVYGGSVTYKVIIGVLDYFFRAGSNPLA 267  
Qy 299 VWEQFTVIGRPAEMPYMAFGQCRGYHDVYELQSVAGYAKAKIPELVMMTDDIDYMD 358  
Db 268 VVOQTQLIAPAPMPYWSFGHOCRYGYNLSOLERVARAYAKAR:PLEVMMTDDIDYMD 327  
Qy 359 AYKQFTLDVNVFPLDKMKKFVNNLHKQKQYVWILDPGIST---NKTYYEYIRGMKHDFV 415  
Db 328 GFKQFTLDVNVFTAELRPFVDRLHRNAQKYVILDFGIRVDPIDATYGTFRGMQODIF 387  
Qy 416 LKRKGPYLGSGWGPVYFPDFLXPALTFWTDBIKRFLNLLPVDGLWIDMNEISNFISS 475  
Db 388 LKRNGTNFVGNWFGDVFYFPDFMEPAAEFWAREISLFRRTIPVDGLWIDMNEISNFIN- 446  
Qy 476 PPIFGSTLDNPPYKINNSGVMLPIINKTIPPTAMHYGDIPEYNVHNLFGYLEARVTRAAL 535  
Db 447 -PEPNALDDPPYINNDGCRPINNTKTRVPLAVHYGGVTEYEHNLFGLEARATGRGV 505  
Qy 536 IKLTEKRPFLVSRTPSGSGKYTAHWTGDNAATWDLVYISPSMLDFGLFGIPMVGADIC 595  
Db 506 LRDGRPPFLVSRSTFVSGSRYTAYMTGDNAATWGLDRLYSINTMLSGLFMPMIGADIC 565  
Qy 596 GFLGNTTEELCRRIOLGAFYPPSRDSSLGTTTQELYRWESVAASARKVLGLRYTLPPY 655  
Db 566 GFNGNTTEELCGRIOLGAFYPPSRDSSAIFTVRELYLWPSVAASGRKALGLRYQLLPY 625  
Qy 656 FYTLMEYLAQLNGIPIARLPFFSPDDIKTYGISQFLLGKGMVSPVLKPGVSVTAFFP 715  
Db 626 FYTLMEYLAHMTGAPIARLPFFSYPHDVATYGVDRQFLLGRGLVSPVLEPGTIVDAYFP 685  
Qy 716 RGNWFDLFDYTRSVTASTGRVYVTLASPPDHNVHIOEGNILAMOGKAMTTOARKTFFHL 775  
Db 686 AGRWYRLYDYSLAVATRGHVRLPAPADTVNHLTGTTILPQSSALTTSRARRTAFFH 745  
Qy 776 LVNWDGCGASFGELFLDDGVETVGNVRGKWTYVKE--IAASAKQTCIITSDVVSSEFAV 833  
Db 746 LVLAEDGTASGYLFDODGDSPEYG-RRDSMSVRFNYKIPNNKGAIKVKSEVWHNSYAQ 804  
Qy 834 SQKWVIDKVTILGRKGTINGYTVRTGAVTRKGKSKLXSTPDRKGFEITVAEISGLNLL 893  
Db 805 SRTLVIKVVLMGHRSPAAPKLLTVHNSAEVASSAGTRYQNAAGLGVAHIGGLSLV 864  
Qy 894 LGRSEFKL 900  
Db 865 VGESEFEL 871

## RESULT 2

US-09-437-054A-8  
; Sequence 8, Application US/09437054A  
; Patent No. 6316698  
; GENERAL INFORMATION:  
; APPLICANT: Allen, Stephen M.  
; TITLE OF INVENTION: Plant Alpha-Glucosidase II Homologs  
; FILE REFERENCE: B61273 US NA  
; CURRENT APPLICATION NUMBER: US/09/437.054A  
; CURRENT FILING DATE: 2001-05-14  
; PRIOR APPLICATION NUMBER: 60/107,909  
; PRIOR FILING DATE: 1998-No. 6316698member-10  
; NUMBER OF SEQ ID NOS: 19

RESULT 3

; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 8  
; LENGTH: 914  
; TYPE: PRT  
; ORGANISM: Glycine max  
US-09-437-054A-8  
  
Query Match 17.4%; Score 831; DB 4; Length 914;  
Best Local Similarity 28.7%; Pred. No. 1e-68;  
Matches 242; Conservative 133; Mismatches 348; Indels 120; Gaps 27;  
  
Qy 75 LSITASLESNDRLVRITD-----AKHRRWEIIPDNILHRHQPPPPPHSLSSLYRTLSS 129  
Db 80 LLTLTSVYQGIILKIDEDPSLSPKKRPEVDVIVSEFPSTKMLWPKISVSENGLSS 139  
Qy 130 P-TTARRKILLSHPNSDLTFSLINTTPEGFTISRKSTHDLVFD-----ATDPNPNMTF 182  
Db 140 VYLSDGHSAVLADHPFEL---FIRDSGDRVISLNSHD-LFDFEQLKHKSDDNWEEQF 195  
Qy 183 LIFIPD-----OYLHLTSSLPGTRAH:YGLGEHS-----KPTPOLAHNQTULTMRAADIPS 231  
Db 196 RSHTDRRYPGPQISIFDVSFYGA-DFYTGIPERAASLALKPT-----RGNVDE 243  
Qy 232 SNP-----DV-----NLYGSHPPFYMDVRSSPVAGSTHGVLLNNSGMDVE----- 271  
Db 244 SEPYRLFNLDVPEYIHDSPFGLYGSIPEMV---SHGKARGSSGPFMLNAAEQIDVLAPG 300  
Qy 272 -----YTGNRIT--YKVIIGIIDLYFFAGSPGVVQFTRVIGRPAEMPYMAFGQ 321  
Db 301 WDAESGIALPISRITDTFMNSEAGVVDAFFFGPNPKDVLRYQVTAVTGTPAMQOLFESIAFH 360  
Qy 322 QCRYGVDHYEIQSVVAGYAKAKIPELVMMTDDIDYMDAYKQFTLDPVNFPLDKMKKFVNN 381  
Db 361 QCRWNYROEEDVEHYDVSKEFDELDIPYDVLWLDIEHTDGKRYFTWDRALFP--HPSEMORK 418  
Qy 382 LHKNGQYVILDDPGISTNKTYYETVIRGMKHDVFLK-RNGKPYLGSGWGPVYFPDFLKP 440  
Db 419 LASKGRHMVTIYVDPHIKRDENFHLKEASQKYYVKDASGNDFDGWCWPGSSSYDPTLNP 478  
Qy 441 SALTFTWDEI--KRFLNLLPVDGLWIDMNEISNFISSPPIFGSTLDNPPYKINNSGVMLP 498  
Db 479 EIRSWMAKFSQSTEGSTPSLYIWNDMNEFSVF-NGPEV----- 517  
Qy 499 IINKTIPPTAMHYGDIPEYNVHNLFGYLEARVTRAALIKLTE--KRPFVLGRSTFSGSK 556  
Db 518 ----TMPRDVTHYGGVEHRELHNA:GYYPHMATANGLLKRGEENDRPFVLSRALPAGSQ 573  
Qy 557 YTAHTGTGNAATWDLVYISPSMLDFGLFGIPMWGADICGFLGNTTEELCRRIWIOLGAFY 616  
Db 574 YGAVMTGDN'TADWDHLRYSIPMVLTLGLTGMFSFGADIGGFFGNPEPELLVRWYQLGAYY 633  
Qy 617 PFSRDHSSLGTTTQELYRW-ESVAASARKVLGLRYTLTPYFVTLMEYLAQLNGIPIARPLF 675  
Db 634 PFFRAAHHDTKRRPFWLFGERNTELIKDAIHRVALLPYFTTLPREANTTGVVVRPLW 693  
Qy 676 FSPDDDIKTYGISQFLLGKGMVSPVLKPGVSVTAFFP-RGNWFDLFD---YTRSVTA 731  
Db 694 MEFPSEATFSDNETFMVSGSSILVQGIYTERAKHASVYLPCKQSWYDLRTGAVYKGVTH 753  
Qy 732 STGRVYVTLASPPDHNVHIOEGNILAMOG--KAMTTOARKTTPFHLVVMDCGASFGEL 789  
Db 754 K-----LEVTEESIPAFORAGTIIARKDRFRSSSTOMA-NDPYTLVVALNASSQAAGEL 806  
Qy 790 FLDDGVETVGNVRCKWTFVVFIAASAKQTCIITSDVVSSEFAVSKWVIDKVTILGRK 849  
Db 807 YIDDD--SSFNPLQGGYTHRRPFIHNGKLTSLDAPASSSKGRYPDSDAFIERIILLGAP 864  
Qy 850 GTK 852  
Db 865 SSK 867





RESULT 5  
US-08-897-843A-1  
; Sequence 1, Application US/08897843A  
; Patent No. 6514493  
; GENERAL INFORMATION:  
; APPLICANT: DeLeo, Albert B.; Loftus, Douglas; Appella, Ettore  
; TITLE OF INVENTION: CDNA CLONE FOR MURINE TUMOR  
; TITLE OF INVENTION: REJECTION ANTIGEN GP110 AND TUMOR PEPTIDE VACCINE  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Diane R. Meyers  
; STREET: 600 Grant Street, 42nd Floor  
; CITY: Pittsburgh  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 15219  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/897,843A  
; FILING DATE:  
; CLASSIFICATION: 424  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 938 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE:  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Mus musculus  
; DEVELOPMENTAL STAGE: Embryo  
; TISSUE TYPE: Embryo  
; CELL LINE: NIH 3T3  
US-08-897-843A-1

Query Match 13.5%; Score 643.5; DB 4; Length 938;  
Best Local Similarity 25.3%; Pred. No. 4,4e-51;  
Matches 247; Conservative 137; Mismatches 361; Indels 231; Gaps 42;  
Qy 52 RRSALTAPQGVKNSVVGPD-----QLSITASLE-----SNDRLRVITDAKHR-2W 99  
Db 50 RPSLPSPSLDTLQL-GPDALTVLHIEVTKVLLVLELOGLQKNTIRIDSEAPRY 108  
Qy 100 EIPONILHRHQPPLPPPHSLSSLYRTLSSPTNRRKILLSHPNS-DLTF-----LIN 152  
Db 109 RVPD-----VLVADPTARLSVSGRDNDSVELTVAEGPYKILT 147  
Qy 153 TTFP-----GFTISRKSTHDLF-----DAPN-DPTN?- 179  
Db 148 AQPFRLDLEDRSLLSVNARGLMAFEHQRAPRVPQSKDPAEGNGAQPEATCDGDKE 207  
Qy 180 -----NPLFLIDQ-----YLHTSLPGTRAHYGLGEHS-----KPT 213  
Db 208 ETQKAEKDEFGAWEETFKTHSDSKPYGTSVGLDFSLPGME-HVYGIPEHADSRLKVT 266  
Qy 214 FOLAHNQTWTWRAADIPSSNPVNLGS-----HPFYMVDR-----SSPVA 254  
Db 267 EGGEFYLRLYNDVQYELNPN-MALYGSVPVLLAHSFHRDLGIFWLNAETWWDISSNTA 325  
Qy 255 GST-HGVLL--LNSGMDVEVTHRIYTKVLCIGIIDLYFFAGSPGQVVEQFTRVGRPA 311  
Db 326 KTLFGKMLDYLGSG---ETPQTDIRMMSSGIDVFLMLGPSVDFVFRQYASLTGTOA 382  
Qy 312 PMPYWAFOQCRGYHVDVYELQSVAGYAKAKIPLEVMMTDIDYMDAYKDFTLDPVNF 371  
Db 383 LPPLFSLGYHOSRWNYRDEADVLEVDQGFDDHNMPCDVIWLDIEHADGKRYFTWTTRFP 442

Qy 372 LDQMKKFNVLHKGQKVVWILDPGISTNKTYETIRGMKHDVFLK-RNGKPYLGSWPG 430  
Db 443 --OPLNMLEHLDKRRNVVAIVDPHIKYVDSGYRVEELRNHGLYKTRDGDSDYEGWCWPG 500  
Qy 431 PVYFPDLKPSALTFTWDEIKRFLNLLPVDG-----LWIDMNEISNFISSPPIPGS 481  
Db 501 SASYPDFINPRBALMS-----NMFSDNYEGSAPNLYVNDMNEPSVF-NGPEV--- 549  
Qy 482 TLNDPPYKINNSGVMLPIINTIPTTAMHYGDIPEYVNHNLFGYLEARVTRAAALIKUTE- 540  
Db 550 -----TMLKDAVHYGCGWEHRDIHNIYGLYVHMATADGLIQSGG 588  
Qy 541 -KRPFVLSRSTFGSGKYTAHTWGDNAATNDLVYSIPSLMDFGLGIPMYGADICGLG 599  
Db 589 IERPFVLSRAFFSGORFGAVWTGNTAEWDHLKISIPMCLSLALVGLSFCGADVGGFFK 648  
Qy 600 NTEELCRRWQLGAFYPPSRDHSLSGTTYOELYRW-----ESVAASA-----RKVLGL 648  
Db 649 NPEPELLVRWYQMGAYGPFRAHAT-----WTLGGESMAVSVSIPRCNPRCLVPA 698  
Qy 649 RYTLPPYFTLMEYEAQLNGIPIAPRLPFSFDDIKTYGISQFLGKGVWSPVLKPGV- 707  
Db 699 IFFAALLVYPLLSSQGRVSCHEAPL-VQYPEDMSTFESIEDQFMLGDALLIHPVSDAGAH 757  
Qy 708 --VSVTAYFPR-GNWFDLFVTRSVTASTGRY-----VTLSPDPDHINVHQEGNILAMQ 759  
Db 758 GORSICLAKKRCGMTF-----RAIRSIMGRPCICPVTLSIP-----VFQGGGTIVPR 806  
Qy 760 GKAM--TTQAAARKTFPHLLVVMSCGASFGBLFDDGVETMGNVGRKWTFKVFIASAK 817  
Db 807 WWRVRRSSDCMKDDPITLFLVALSPQTAQGBLFDGHTFNY-QTRHEFLRRF---SFS 862  
Qy 818 QTCILITSDVVSGEFAVSOKWIDKVTILGLRKGTINKGYTVRTGAVTRKGD-KSKLKSTP 876  
Db 863 GSTUVSSADPKHLETFIW-IERVIMGAKPAAV-----VLQTKGSPESRSLSPQH 913  
Qy 877 DRKGEFIVAEISGLNL 892  
Db 914 DPETSVLILRKPGYSV 929  
RESULT 6  
US-08-633-770A-2  
; Sequence 2, Application US/08633770A  
; Patent No. 5908760  
; GENERAL INFORMATION:  
; APPLICANT: Bojsen, Kirsten  
; APPLICANT: Yu, Shukun  
; APPLICANT: Kragh, Karsten  
; APPLICANT: Christensen, Tove  
; APPLICANT: Marcussen, Jan  
; TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM A FUNGUS, ITS  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson & Bear  
; STREET: 620 Newport Center Drive 16th Floor  
; CITY: Newport Beach  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/633,770A  
; FILING DATE: July 8, 1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP94/03398  
; FILING DATE: OCT-15-1994



```
Db 202 -----HKLKDTVLIVKPG-----HGEYVGVGEMGQIQ 230
QY 209 -HSKPTF-----QLAHNQTULTMAADI PSSNPDVNLVYSGHPFVMDVRSSEVAGSTHGV 260
Db 231 FKKEFTFMYFNFDNMVQOQVYAQALD--SREP--LYHSDPFLVDVNSPEHNTAT 285
QY 261 LLNSNGMDVEYTGNRITKVIJG---GIDYFFAGSPGQVQFTRVIGRPAKMPYWA 317
Db 286 FIDNYSQIAIDFGKNSGYIKLTRYGGIDCYGISADTVPEIVRLYTGIVGRSKLKPRYI 345
QY 318 FGQOQRCYGHVYELQSVVACAKIPLVEMWTDIDYMDAYKDFLDPVNFPLDKKK 377
Db 346 LGAHAQACYQOESDLYSVQOYRCKPFLDGIHVDVQDQFRTFTTNPHFTF--NPKE 403
QY 378 FVNNLHKNQKQVWIIDPGISTNKT---YETVIRGMKHDFVLFK-----418
Db 404 MFTNLRRNGIKSTWITPVISINNREGGYSTLLEGVDKKYFIMDDRYTEGSGNAKOVRY 463
QY 419 -----NGK-----PYLGSVMPG-----PVFPDFLK 439
Db 464 MYGGGNKVEVDPNVNGRPFDFKNDYFPANFNSKQYYPHGGVSYGNGSAGFYDNLN 523
QY 440 PSALFTWTDIEKRFNLPLVDGLMDMNEISNFISSPPIPGSTLDPNPPYKINNSGVMLPI 499
Db 524 KEVRIWGMQYK-YLFDMLGLEFVQDM-----TTPAHTSYGDMKGLPRLILVTSDSV 575
QY 500 INKTIPPTAMHVGDIPYNNVNL-----FGYLBARTVRAALIKLTKRPFVLSRSTFGSG 555
Db 576 TWASEKKLAETWALYSYNLKRATWHGLSRLESR-----KNKRNFIILGRGSYAGAY 626
QY 556 KYTAHWTGNATWDLVYSISMLDPLGLFGIPMYGADICGF-----LGNTEBELC---606
Db 627 RFAGLTGDNASNWEFWKISVSQVLSLGLNGVCIAGSDTGGFEPYRDANGVEEYKCSPEL 666
QY 607 -----RRMIQLGAFYP-FSRDHSLTGITTYOELYRWESVAASAR 643
Db 687 L-RWYTGSEFLPLWRNHYVKKORKNFQEPYSYPKELETHPELA---DQAWLYKSVLEICR 743
QY 644 KVLGLRYTLPPYILMYEAQLNGIPIARPLFPSPDDIKYG-----ISSQFLLGKGVX 698
Db 744 YVVELYSILQLLYDCMFQNVGDGMPITRSMLLTDTEDTTFNESQKFLDQYVAGDDIL 803
QY 699 VSPVLX-----PGWVSVAIYFP-RGNWF-----DLFDVTRSVTA 731
Db 804 VAPILHSRKEIFG-ENRDVYLPYHTWYPSNLRPNWDQGVALGNPVEGGSVINYARIV- 861
QY 732 STGRYVTLSPDPH-----INVHQEGNIL-AMQOKANTQ-AARKTPFHLLVVMSCDG 783
Db 862 -----APEDYNLFHSPVYVYREGAIIPOIEVRQWQGGANRIKFIYIP-----906
QY 784 ASFGE-----LFLDDGV 795
Db 907 ---GKQKEYCTYLDGV 920

RESULT 8
US-09-275-608-3
; Sequence 3, Application US/09275608
; Patent No. 6541237
; GENERAL INFORMATION:
; APPLICANT: YU, Shukun
; APPLICANT: JSEN, Kirsten
; APPLICANT: MARCUSSEN, Jan
; TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE AND
; TITLE OF INVENTION: ITS USE IN THE PRODUCTION OF 1,5-AHYDROFRUC
; TITLE OF INVENTION: TOSE
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
```

```
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/275,608
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: PCT/EP95/02172
; APPLICATION NUMBER: 06-JUN-1995
; FILING DATE: 08/836,156
; APPLICATION NUMBER: PCT/EP94/03397
; FILING DATE: 15-APR-1997
; APPLICATION NUMBER: PCT/EP94/03397
; FILING DATE: 15-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Altman, Daniel E
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: DY09,001C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714-760-0404
; TELEFAX: 714-760-9502
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1092 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-275-608-3

Query Match 7.8%; Score 371.5; DB 4; Length 1092;
Best Local Similarity 21.6%; Pred. No. 2,1e-25;
Matches 196; Conservative 132; Mismatches 380; Indels 199; Gaps 36;

QY 80 SLESNDRLVRITDAKRRWEIPDNILHRHQPDPHPHSLSSLYRTLLSSPTTRRKILL 139
Db 120 SVSNHNPVQIQVTSYNNNSFRIFN-----PDGPIRDSRGPIKQOJTWIRNQELA 171
QY 140 SHPNSDLTESL-----INTTFFGTISRKSTHVLFDATPDPTNPNT-----FL 183
Db 172 QGCNPNMSPSPGFLSFETKOLNVIIYGCNVRVTKOGYLVMEDECNSSQSGNKCRGL 231
QY 184 IFIDQVHLHTSSLPGTRAH-----IYGLGE-----HSKPTPOLAHN-----219
Db 232 MYVDRLYGNIAIASVQTNFHKDTSRNEKFYGAGEVNCRYEEQKAPTYVLSRGLAMNTYN 291
QY 220 -QTLTMRAAD-IPSSNPD-----VNLVGSHP-----FYMDVRSSPVAG 255
Db 292 YNMLNYPQDVVPFGYPDPHPNYIIPMYAAFWLVQGCAGTSKQSYGFWFMDVNSQSYMN 351
QY 256 STHGVLNLSNGMDVEYTGNRITKVIJGIIIDYFFAGSPG--QWVEQFTRVIG-----308
Db 352 T--GDTAMNCGQENLAYMGAQY-----GPFQHFVYGGDGLDVEDVKAFLQSGKEFED 403
QY 309 -----RPAPMPYNAWGFQOCRYGHDVYE-----LSVVVAGYAKAKIPLVEMWT 352
Db 404 KKLNRKSNMPPKYPYVFGFQGVGVALSLKLQNLPAGENNISVQEIIVEGYQDNDFPEGLAV 463
QY 353 DIDYMDAYKDFLDPVNFPLDKN-----KKFVNNLHKNQ---QKTVWIIDPGISTNKK 401
Db 464 DVDMDQDLVRFTTKPEYWSANMVGEGDPNNRSVEFAHADRGLVCQTVTCFLRVDNSGK 523
QY 402 TYETVIRGMKHDFVLFKRN-----GKPYLGSVMPG-----PVFPDFLKPSA 442
Db 524 PYEVNQTLEKQLYTKNDLSLNTDFGTTSDGPDGDAYIGHLDYGGGVCEDAIFPDWGRPDV 583
QY 443 LTFWTDIEIKRFLNLLPVDGLWIDM-----NEISNFISPPPIPGSTLIDNPP-----487
```

584 AQWGENYKFLS-IGLDFVWQDMVTPAMPHRLGDAVNKNSSGSSAPGMPNENDPSNGRY 642  
QY 488 -YKINSGVMLPIINKTIPPTAMHYGDIPEYVNVHNLFGYLEARVTRAALI-----KLTE- 540  
Db 643 NKSYPHQVLV-----TDMRYGAEYREPVMVSQBNHAYTLCESTRREGIVGNADSLTKF 697  
QY 541 KRPFVLSRSTFSGSGKYTAHTWTGNAATWMDLVYSIPSMDFGLFGIPMWGADICGFL-- 598  
Db 698 RRSYIISRGYIGNQHFGGMVWGDNSATESYLOWMLANIINNMWCLPLVGSIDIGFTQY 757  
QY 599 ----GNTTEELCRWILQAGAFYPPSRDH-----SSLGTYOELYRWESVAASARKVLG 647  
Db 758 NDAGDPTPELDMRVFVQAGCLLPFRHRYDRWIESKKGKXYOELYMPQOKDTLKKFVE 817  
QY 648 LRYTLPLFYFTLMYEAQLNGIPI--ARPLFFSPDDIKTY-GISSOFL-----GKGWMS 700  
Db 818 FRYRQEVLYTAMYNQATTGEPILKAAPM---YNDVNVYKSDHFLGCGHDGYRLCA 874  
QY 701 PVLKPGVSVTAYFP-RGNWFDL---PDYTRSVTASTGRVVTI---SAPPDHINVHIQEG 753  
Db 875 PVVRENATSREVLPVYSKWFKFGPDF-TKPLENEIQGGOTLYNYAAPLNDSPFVREG 933  
QY 754 NILAMO-----GKAMTTOARKTPPHLLVWMSDCGASFGELPLDDGVEVTMCVNRGKWT 807  
Db 934 TILPTRLTGWNKSINTY-TDNDPLVFEPLPLENNQAHGLFTHDDGGVTTNAEDFGKYS 992  
QY 808 FVKFIAA 814  
Db 993 VISVKAA 999

RESULT 9

US-08-633-768A-1  
Sequence 1, Application US/08633768A  
Patent No. 6013504  
GENERAL INFORMATION:  
APPLICANT: YU, SHUKUN  
APPLICANT: BOUSEN, KIRSTEN  
APPLICANT: KRAGH, KARSTEN  
APPLICANT: BOJKO, MAJA  
APPLICANT: NIELSEN, JOHN  
APPLICANT: MARCUSSEN, JAN  
TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM  
TITLE OF INVENTION: A FUNGUS INFECTED ALGAE, ITS PURIFICATION  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/633,768A  
FILING DATE: 02-JUL-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 9321301.5  
FILING DATE: 15-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Altman, Daniel E  
REGISTRATION NUMBER: 34,115  
REFERENCE/DOCKET NUMBER: DY0U7.001AEC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 714-760-0404  
TELEFAX: 714-760-9502  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1088 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-633-768A-1  
Query Match  
Best Local Similarity 23.0%; Pred. No. 1.3e-23; Length 1088;  
Matches 200; Conservative 107; Mismatches 347; Indels 215; Gaps 39;  
QY 146 LTF-----SLINTTFFGFTISRKSTHDVLF-----ATPDTPNPTNLTIFIDQ-YLHTSS 195  
Db 186 LTFETKDLVIIYGNFKTRVTRKSDGKVIEMENDEVTASSGNKCRGLMFVDRLYGNALAS 245  
QY 196 LPTG-----RAHYILGEGHS---KPTF-----OLAHNQTLTWRAADIP 230  
Db 246 VKNFRNDVAKOEGFYGAGEVNCYQDTYILERTGLIAMTNYNOLNLYNOMDLRPPPHDG 305  
QY 231 SSNPD--VNLVGSHP-----FYMDVRSSPVAGSTHGVLLNLSNGMDVE 271  
Db 306 ALNPDYIPMYAAPLLVINCAGTSBOYSYGFMDNVVSQSYMNT--GDTTWSGQEDLA 363  
QY 272 YTGNRITYKVIIGIIDLYFFAGSPGQ--VVEQTRVIG-----RPAPMPYWAFG 319  
Db 364 YMAQY-----GPFQHFVYGAGGMECVVTAFLSQGKEFEFNOVLNKRWSMPKVFVG 417  
QY 320 FQOCRYGYHDV-----YELQSVAGYAKAPILEVMTDIDYMDAYKDFTL--- 365  
Db 418 PFQGVFGTSSLRAHPAGENNISVEEIVEGYQNNPFEGLEAVDVMQDLRVFTTKGE 477  
QY 366 -----DPVN-----PPLDK-----MKFVNHLHKGQKYVILDPGISTNKT 402  
Db 478 FWTANRVCTGSDPNRSVFEMAHDKGLVCQTNITCFLEN-DNEGQDVEV-----NQLTRE 531  
QY 403 YETIIR-----GMKHDVFLKRNKGPYLGWVG-----PVYPDFLKPSALTFTWD 448  
Db 532 RQLYTKNDLSLTGDTFGMTDD-----GPSDAYIHLIDYGGVECDALFPDWCGRPDVAEWGN 587  
QY 449 EIKRFLNLLPVDGLWIDMNEISNFISSPPIPGSLDNPYPKYNNSGVMPLIN----- 501  
Db 588 NYKLLFS-IGLDFVWQDMT-----VPAMM-----PHKIGDDINVKPGDNPWADDP 632  
QY 502 -----KTIPP-----TAMHY-----GDIPEYVNVHNLFGYLEARVTRAALI-----KLTE-K 541  
Db 633 SNGQVNWKTYPQVLVTDNRVENHGREPMVTQRIHAYTLCESTRKEGIVENADTLTKFR 692  
QY 542 RPFVLSRSTFSGSGKYTAHTWTGNAATWMDLVYSIPSMDFGLFGIPMWGADICGFLGNT 601  
Db 693 RSYIISRGYIGNQHFGGMVWGDNSITTSNYIGMMIANNINNMWCLPLVGSIDIGFTSYD 752  
QY 602 TE-----ELCRRWIOLGAFYPPSRDH-----SSLGTYOELYRWESVAASARKVL 646  
Db 753 NENQRTCTGDLWRYVQAGCLLPFRHRYDRWIESKKGKXYOELYMPNEMDTLTKFV 812  
QY 647 GLRYTLPLFYFTLMYEAQLNGIPIARPLFFSPDDIKTYGISSOFL-----GKGWMSVP 702  
Db 813 EFRYRWGEVLYTAMYNQAAAPGKPIKAASMYNDSNVRRANQNDHFLGCGHDGYRLCAPV 872  
QY 703 LKPGVSVTAYFP-RGNWFDL---PDYTRSVTASTG--RYTFLSAPPDHINVHIQENIL 756  
Db 873 VWENSTERELYPVLTOQWYKFGPDFTKPLEGANMGGDRIYNYVPVPOSESIFVREGAIL 932  
QY 757 AM-----QKAMTTOARKTPPHLLVWMSDCGASFGELFDDGVEVTMGVNRKWTVPVK 810  
Db 933 PTRYTLGENKSLNTY-TDEDPLVFEVFPPLGNRPNADGMCYLDGCGVTTNAEDNGKFSVVK 991  
QY 811 FIAA--SAKQTCIITSD-----VVSGEFAV 833  
Db 992 VAAEQDGGTETITFTNDCEYEVFGGPFVV 1020

QY	469	ISNFSSPPICSTLDNBPYKINNSGVMLPIINKTIPPTAMHY----	GDIPVNVNLFQY 529
DB	620	VGDADVTRSPYGPWNEPNDP---SNGRYNWKSHPOVLVDHRYENHGREPMTORNMAY 678	
QY	526	LEARVTR-----AALIKLUTE-KRPFVLSRSTFSGSKYTAHTMTGDNAATMNDLVYSIPSM 579	
DB	677	TLCESTRXEGIVANADTLTKFRSYYIISRGVYIGNOHFGWVGDNSSQRYLQWMIANI 736	
QY	580	LDFGLFGIPMGADICGFLGNTTEELCR-----RWOLGAFYFSPSRDH-----SSLGT 627	
DB	737	VNNMSCLPLVSGDIDGGFTSYDGRNVCPGDLWVRVQAGCLLPFRNHYGRVLVKGQEGK 796	
QY	628	TYOELYRMESVAASARKVGLGRYTLPLFYFYLMTYEAQLNGIPIARPLFPSPDDIKTYGI 687	
DB	797	YYOELVYKDEMATLKRPIEPRYRQEVLYTAMYQNAAPGKFIKAASM-YDNRNVRGA 855	
QY	688	-SSQFL-----GKVMVSVPLKPGVVSVAFFP-RGMWFDL---FDYTRSVTASTGRVV- 737	
DB	856	QDDHFLGLGHGDRYILCAPVWMENTSRDLVLUVTWKYKFGPDYDTKRLDSALDGGQMI 915	
QY	738	-TISAPPDHINVHIOEGNILAMQ-----GKAMTTOAARKTFPHLLVVMSCGASFGELF 790	
DB	916	KNYSVPQSDSPFVREGAILPTRYTLDSGNSKSMNTY-TDKPLVFEVPLGNRADMCMY 974	
QY	791	LDDGVETVMVNRGKWTFKVFTAAAKOTCIITSVSGEFAV-SQKVID---KVTLIG 846	
DB	975	LDDGTTTDAEDHGKPSVINVEALRKGVTTTI-----KFAIDTYQYVFDGPFYVIRN 1027	
QY	847	LRKGTXYNGTYVTRGA 862	
DB	1028	LTTASKIN---VSSGA 1040	
<p>RESULT 11</p> <p>US-09-275-608-4</p> <p>Sequence 4, Application US/09275608</p> <p>Patent No. 6541237</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: YU Shukun</p> <p>APPLICANT: JSEN, Kirsten</p> <p>APPLICANT: MARCUSSEN, Jan</p> <p>TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE AND</p> <p>TITLE OF INVENTION: ITS USE IN THE PRODUCTION OF 1,5-ANHYDROFRUC</p> <p>TITLE OF INVENTION: TOSE</p> <p>NUMBER OF SEQUENCES: 21</p> <p>CORRESPONDENCE ADDRESS:</p> <p>ADDRESSEE: Knobbe, Martens, Olson &amp; Bear</p> <p>STREET: 620 Newport Center Drive 16th Floor</p> <p>CITY: Newport Beach</p> <p>STATE: CA</p> <p>COUNTRY: U.S.A.</p> <p>ZIP: 92660</p> <p>COMPUTER READABLE FORM:</p> <p>MEDIUM TYPE: Diskette</p> <p>COMPUTER: IBM Compatible</p> <p>OPERATING SYSTEM: Windows</p> <p>SOFTWARE: FastSeq for Windows Version 2.0b</p> <p>CURRENT APPLICATION DATA:</p> <p>APPLICATION NUMBER: US/09/275,608</p> <p>FILING DATE:</p> <p>CLASSIFICATION:</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: PCT/EP95/02172</p> <p>FILING DATE: 06-JUN-1995</p> <p>APPLICATION NUMBER: 08/836,156</p> <p>FILING DATE: 15-APR-1997</p> <p>APPLICATION NUMBER: PCT/EP94/03397</p> <p>FILING DATE: 15-OCT-1994</p> <p>ATTORNEY/AGENT INFORMATION:</p> <p>NAME: Altman, Daniel E</p> <p>REGISTRATION NUMBER: 34,115</p> <p>REFERENCE/DOCKET NUMBER: DY009.001C1</p>			

## TELECOMMUNICATION INFORMATION:

TELEPHONE: 714-760-0404  
TELEFAX: 714-760-9502  
TELEX:  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 570 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-275-608-4

Query Match 7.0%; Score 335; DB 4; Length 570;  
Best Local Similarity 24.4%; Pred. No. 1.8e-22;  
Matches 122; Conservative 71; Mismatches 188; Indels 118; Gaps 18;  
QY 269 DVEYTGRTYKVIIGGIIDLYFF--AGSPSQVVEQFTRVIG-----RPARMPYM 316  
DB 76 DLAHYGAQC-----GPFDOHFVVEAGDGLVEDVVTAFSYLQGEYENQGLNIRSNMPPKY 129  
QY 317 AFGQOQRYGHDV-----YELQSVVAGYAKAKIPILEVMWTDIDYMDAYKDFTL 365  
DB 130 VFGFQGVFGATSLRLNLPAGENNVSLEIVEGYQNVPPFGLAVDVMQDDLEVF 189  
QY 366 DPNVNFPLDKMKKFVNNLHKNGQKVVWILDPG1--STNKTVETVIRGMKHGVFLKRNKGPY 423  
DB 190 RPAFTANKYGEQDPNNKSVFEWA--HDRGLVCQTNV-----CFLKNEKNPY 236  
QY 424 -----LGSVMWGP-----VYPPDFLKFSALT 444  
DB 237 EVNQSLREKQLYTKDSLNDIFGTTDPGFSDAYIGHLDYGGVGVEDALFPDWGRPDVAQ 296  
QY 445 FWTDEIKRFLNLLPVQGLWDM-----NEISNFI---SSPPIPGSTLDNPPYKINNSG 494  
DB 297 WMGDNYKKLFS-IGLDFVQDMVTVPAMPHRLGDPVGTNSETAGPNDKDP---SNGR 352  
QY 495 VMLPIINKTIPPTAMHYGDI---PEYNVHNLFGYLEARVTRAALI-----KLTE-KRPV 545  
DB 353 YNWKSYHPQVLVDMRYDDYGRDPIVQTNLHAYTLCESTRREGVGNADSLTKFRSVI 412  
QY 546 LSRSTFGSGSKYTAHTGDNNAATWNLVYIPSMDFGLGPIWAGADIGGFLGN-----600  
DB 413 ISRGYIGNOHFGGMMYVGDNSTEDYLAMVIVINNMNSGVPLVGDSDIGGFTEDHKRNP 472  
QY 601 TTELCRRWTLGAFYFSPSRDH-----SSLGTTTQELYRWESVAASARKVLGLRYTL 653  
DB 473 CTPDLMRRFVQAGCLLEWFRNHYDRWIESKKHKNYQELMYRDLDAURSFVELRYRW 532  
QY 654 PYFTYLMYEAQLNGIPIAR 672  
DB 533 EVLYTMYQNALNGKPIIK 551

## RESULT 12

US-09-634-238-331  
Sequence 331, Application US/09634238  
Patent No. 6544772  
GENERAL INFORMATION:  
APPLICANT: Glenn, Matthew  
APPLICANT: Havukkala, Ilkka J.  
APPLICANT: Bloksberg, Leonard, N.  
APPLICANT: Lubbers, Mark W.  
APPLICANT: Dekker, James  
APPLICANT: Christensson, Anna C.  
APPLICANT: Holland, Ross  
APPLICANT: O'Toole, Paul W.  
APPLICANT: Reid, Julian R.  
APPLICANT: Coolbear, Timothy  
TITLE OF INVENTION: Polynucleotides, materials incorporating  
them and methods for using them.  
FILE REFERENCE: 11000.1043U1  
CURRENT APPLICATION NUMBER: US/09/634,238

CURRENT FILING DATE: 2000-08-08  
NUMBER OF SEQ ID NOS: 422  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 331  
LENGTH: 390  
TYPE: PRT

ORGANISM: Lactobacillus rhamnosus

US-09-634-238-331

Query Match 4.2%; Score 199; DB 4; Length 390;  
Best Local Similarity 22.8%; Pred. No. 5.8e-10;  
Matches 75; Conservative 53; Mismatches 159; Indels 42; Gaps 11;  
QY 417 KENGKPYLGSVWPGVYFPDFLKPSSALTFTWDEIKRFLNLLP-----VDGLWIDMNEISNF 472  
DB 77 QKNOMP-LGMFLPDNGGAGYGQDTTLAGNLQNLKSPADYADQHGVAATGLWTOQN-----130  
QY 473 ISSPPIPGSTLDNPPYKINNSGVMLPIINKTIPPTAMHYGDIPEYNVHNLFGYLEARVTR 532  
DB 131 -LSPVDPA---NPKPDDRDFAKEVALGVKALKTDVAWVGSGYSFGLDGL-AKADAMMTQ 184  
QY 533 AALIKLTKTEKPPVLSRSTFGSGSKYTAHTGCD-NAATWNLVYIPSMDFGLGPIWAG 591  
DB 185 ---VKGDSLRFPAITLDGWAGTQRYAGVWTDGDTGGQWMEYIREHIPTIYIGTGLSGQPYVG 241  
QY 592 ADICGFLGNTTEELCRRWIQLGAFYFSPSRDHSSLSGTTTQELYRWESVAASARKVLGLRYT 651  
DB 242 SDMGIFGGGNPIVNTDRDFQWKAFTPIQLNDGNGANPKTFFDFDQOTTAINRAYNKQKT 301  
QY 652 -LLPYFTYLMYEAQLNGIPIARLFFSPDDIKTYG--ISSQFLLGKGMVSPVLKPGVV 708  
DB 302 MLMFYNVTASAGSVFDGKPMVGRGLFDYPNIPEAYTDLVKVEYLWGDNFLVAPIYQN---358  
QY 709 SVTAYFPRGN-----WFDLF 723  
DB 359 -TAADEKGNDRNGIYLPDKQVWIDYY 385

## RESULT 13

US-09-437-054A-10  
Sequence 10, Application US/09437054A  
Patent No. 6316698  
GENERAL INFORMATION:  
APPLICANT: Allen, Stephen M.  
APPLICANT: Kinney, Anthony J.  
TITLE OF INVENTION: Planc Alpha-Glucosidase II Homologs  
FILE REFERENCE: BB1273 US NA  
CURRENT APPLICATION NUMBER: US/09/437,054A  
CURRENT FILING DATE: 2001-05-14  
PRIOR APPLICATION NUMBER: 60/107,909  
PRIOR FILING DATE: 1998-No. 6316698ember-10  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 10  
LENGTH: 150  
TYPE: PRT  
ORGANISM: Triticum aestivum  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (73)..(74)  
NAME/KEY: UNSURE  
LOCATION: (102)  
NAME/KEY: UNSURE  
LOCATION: (123)  
NAME/KEY: UNSURE  
LOCATION: (141)  
NAME/KEY: UNSURE  
LOCATION: (143)  
US-09-437-054A-10

Query Match 4.0%; Score 190.5; DB 4; Length 150;

Best Local Similarity 29.4%; Pred. No. 7.1e-10;  
Matches 45; Conservative 33; Mismatches 62; Indels 13; Gaps 5;

QY 648 LRYTLLPYFTLMEYAOINGIPIARPLFFSPDDIKTYGSSQFLKGVVSVPLKGV 707  
Db 4 KRYSLPYYISLQFASVTVGVVWPLWLEFPDDKKTNNNGEAMVGSILAQGIYEEGQ 63  
QY 708 VSVTAYFRGN-WFLDFD---YTRSVTASTGRYVTLSPAPPDHNVHIOEGN:LAMQG-K 761  
Db 64 KSVSVYLCXLLWYDLRNGSPKGSVSHK-----LQVSEDSIPXFORSGTIVPRKDRFR 117  
QY 762 AMTQOARKTPHLLVMSDCASFGELFLDDG 794  
Db 118 RILTOXVND-S-YTLVIGLNSWLVKXTYVDDG 149

RESULT 14  
US-08-551-356-2  
; Sequence 2, Application US/08551356  
; Patent No. 5830700  
; GENERAL INFORMATION:  
; APPLICANT: Irani, Meher  
; TITLE OF INVENTION: HYBRID CROSS-LINKING PROTEINS  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ZymoGenetics, Inc.  
; STREET: 4225 Roosevelt Way, N.E.  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98105

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/551,356  
; FILING DATE:  
; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/349,762  
; FILING DATE:  
; APPLICATION NUMBER: US/07/998,272  
; FILING DATE:

; ATTORNEY/AGENT INFORMATION:  
; NAME: Holly, Julie A  
; REGISTRATION NUMBER: 33-246  
; REFERENCE/DOCKET NUMBER: 92-26  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-547-8080 ext 322  
; TELEFAX: 206-548-2329  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2446 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

US-08-551-356-2

Query Match 2.7%; Score 130.5; DB 2; Length 2446;  
Best Local Similarity 18.2%; Pred. No. 0.037;  
Matches 180; Conservative 114; Mismatches 324; Indels 373; Gaps 44;  
QY 125 TLLSPTNRRKILLSHPNSDL---TFLSINTPFPGTISRKSTHD-----VLFDATPD 175  
Db 700 TSTSPVTSNTVGTTPSPPLVAISEVTEITASSFVSVWSASDTSVGERVEYELSEE 759  
QY 176 PNTNTELFIDQYHLHTSS-----LPQTR--AHYGLGKSKPTFOLAHNQTLM 224  
Db 760 GDPE-----QYDLRSTATSNIPOLLPRKRVVNVYQISEDEGQSLILSTQITAP 811  
QY 225 RAADIPS-----SNPDVNLXGSHPPFYNDVRSSPVAGSTHGVLLNSNGMDVE 271  
Db 812 DAPPDPTVDQVDDTSIVVWMSRQAPITG-----YRIVYSPSVEGS-----STELNLP 859

QY 272 YTGNRITYKVI--GGIIDLFFFA-----GSPQGVVEQFTRVIGRPAPMPYWAFFGQOCXY 325  
Db 860 ETANSVTLSDLGQGVQYNIITYAVEENQESTFVVIQOETTGTGPRSDTVP----- 908  
QY 326 GYHDVYELQSVVAGYAKAKIPLVENVMTDIDYMDAYKDFLTD--PVNFPDLKMKKFFVNLH 383  
Db 909 ---SPDLQFVEV---TDVKVTIMWTPPE--SAYTGRVDVIPVNLPGEGHQRL--PIS 957  
QY 384 KNGQKYVWILDEGIS-----TNKTYET----- 405  
Db 958 RNTFAEVTGLSPGVTYFKVFAVSHGRESKPLTAQOTTKLDAPTNLQFVNEDTSVLVRW 1017  
QY 406 -----YIRGMKHDVFLKNGKPYLSVWPGVYFPDFLKPSSALTFTWDEIRKFLNLLPVD 460  
Db 1018 TTPRAQITGYRLTVGLTRGQPROYNVGSVSKYP-----LRNLQAS 1060  
QY 461 GLWIDMNEISNFISSPPI-----PGSTLDPNPKY----- 489  
Db 1061 EYTVSLVAIKNQESPKATGVTTLOPGSSI--PPYNTVETTIVITWTPAPRIGFKLG 1118  
QY 490 -----INNSGVM-----PIINKTI-----PPT 507  
Db 1119 VRPSQGEAPREVTSDSGSIVVSGLTPTGVEYVYTIQVLRHGOERDAPVKNVVTPLSPPT 1178  
QY 508 AMHYGDIPEYV-----HNLFGYLEARVTRAAAL-----IKLTEKRPVLSRSTFS 552  
Db 1179 NLHLEANPDGTLTVSWERSTTPDITGY---RITTPNGQOGNSLEEVVHADQSCTPD 1235  
QY 553 GSGK-----YTAHWTGDNAATWDLVYSPSLMDFGLFGIPMVGADICGLGNTTEE 604  
Db 1236 NLSPLGLEYNVSVTVKDDKESVPISDTIIPVQPQLTDLFSVDI----- 1278  
QY 605 LCRWITQLCAFYPSPDRHSSLTQYELYRWSVAASARKVLGLRYTLPLPYTYLMEYAO 664  
Db 1279 -----TDSSIG-----LRWTPLNST--IIGYRITV-----AA 1305  
QY 665 LMGIPARPLUFFSPDDIKTYGSSQFLKGVVSVPLKGV-----VSVTAYFRGNWFD 721  
Db 1306 GEGIPIFEDFVYS---SVGYVYVVG-----LEPGIDYDISVITLINGE--- 1346  
QY 722 LFDYTRSVTASTG-----RYVTLS-----APDHNHVIQEGNILAMQKAMTQ 766  
Db 1347 --SAPTTLTQQTAVPPTDLRFTNIGPDTRMTVWAPPSSIDL-----TNFLVYSPVKNEE 1400  
QY 767 AARK---TPHLLVMSDCASFGELFLDDGVEYVMTGVN-----RGKWTFKVFI 812  
Db 1401 DVAELISPSDNAVLTN-----LLPGTEYVSVSVSYEQHESTPLGRQK----- 1446  
QY 813 AASAKQTCIITSDVWSGEFASQKRWIDKVTILGLRKGTKINGYTVRTGAVTRKDGSKL 872  
Db 1447 TGLDSPTGIDFSDITANSFTV--HMAIAPATITGYRIHHPEHFSRPREDRVPHSRNSI 1504  
QY 873 KSTPDRKGEFIVAEISGLN-----LLGRE 897  
Db 1505 TLTNLTGTEYVSVVSVIALNGREESPLLIGQQ 1535

RESULT 15  
PCT-US93-12687-2  
; Sequence 2, Application PC/TUS9312687  
; GENERAL INFORMATION:  
; APPLICANT: Irani, Meher H.  
; TITLE OF INVENTION: HYBRID CROSS-LINKING PROTEINS  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ZymoGenetics, Inc.  
; STREET: 4225 Roosevelt Way, N.E.  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98105  
; COMPUTER READABLE FORM:



```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/12687
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/998,271
; FILING DATE: 31-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, Gary E
; REGISTRATION NUMBER: 31-648
; REFERENCE/DOCKET NUMBER: 92-26PC
; TELEPHONE: 206-547-8080 ext 322
; TELEFAX: 206-548-2329
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2446 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US93-12687-2
```

```

Query Match          2.7%; Score 130.5; DB 5; Length 2446;
Best Local Similarity 18.2%; Pred. No. 0.017;
Matches 180; Conservative 114; Mismatches 324; Indels 373; Gaps 44;

QY 125 TLLSPITNRKILLSPNSDL---TFSLINTTPFGFTISRKSTHD-----VLFDAITPD 175
DB 176 PTNWFLEIFIDVHLTSS-----LPCTR--AHYGLGEHSKPTFQLAHNOTLTX 224
DB 700 TSTSTPTSTNVTGETTFFSPPLVATSEVTEITASSFVSWVSASDTVSGFRVEYELSEE 759
QY 176 PTNWFLEIFIDVHLTSS-----LPCTR--AHYGLGEHSKPTFQLAHNOTLTX 224
DB 760 GDEP-----QYLDLPSTATSNIPDLPGRKIYVNVYQISEDGEQSLILSTQTTAP 811
QY 225 RAADIPS-----SNPDVNLVGHSPFYMDVRSSPVAGSTHGVLILLNSNGMOVE 271
DB 812 DAPPDPTVDQVDDTSIVVRWSPQAPITG---YRIVYSPSEGS-----STELNLP 859
QY 272 YTGNRITYKVI--GGIIDLXYFA---GPSQGVVEQFTRVIGRPAMPYWAFFGOOCRY 325
DB 860 ETANSVTLSDLQPGVQYNIITVAYEENQESTPVVIQOETTGTGRSDTVP----- 908
QY 326 GYHDVYELQSVVAGYAKAKIPLVNMWTDIDYMDAYKDFDLD--PVNFPDLMKKFVNVLH 383
DB 909 ---SPRDQGFVEV---TDVKVTIMTTPPE--SAVTGRVDVIPVNLPGHGQRL--PIS 957
QY 384 KNGQKVVVILDPGIS-----TNKTYET----- 405
DB 958 RNTFAEVTGLSPGVITYYKVPFAVSHGRESKPLTAQOTTKLDAPTNLQFVNETDSTVLVRW 1017
QY 406 -----YIRGMKHDFVLKNGKPYLGSVWPGVYFPDFLKPALSALTFTWDE-KRFLNLLPVD 460
DB 1018 TPPRAQITGYRLTVGLTRGQFRQTNVGPVSKYP-----LRNLQPAS 1060
QY 461 GLWIDMNEISNFISSPPI-----PGSTLNDNPPYK----- 489
DB 1061 EYTVSLVAIKGNOESPATGVFTTLQPGSSI--PPYNTVEVTITVITWTPAPRIGFKLG 1118
QY 490 -----INNSGVM-----PIINKTI-----PPT 507
DB 1119 VRPSQGEAPREVTSQSGSIVVSGLTPGVYVYTIQVLRDQGERDAPVNVKVVTPLSPT 1178
QY 508 ANHYGDIPEYV-----HNLFGYLEARVTRAL-----IKLTEKRPFFVLSRSTES 552
DB 1179 NLHLEANPDTGVLTVSWERSTTPTDITGY---RITTPNGQQGNSLEEVHADQSSCTFD 1235
QY 553 GSGK-----YTAHWTCGDNAATWDLVYIPSLMDFGLFIPVGVGADICGFLGNTTEE 604
DB 1236 NLSPGLEYNVSVYTVKDDKESVPISDTIIPVLPQLTDLSPVDI----- 1278
```

Search completed: October 27, 2003, 10:26:47  
Job time : 22.3223 secs

```

QY 605 LCRRWIOLGAFYPPFRDHSLSGTTYOELYRWESVAASARKVLGLRYTLPLPYFTLMEYEQ 664
DB 1279 -----TDSSIG-----LRWTFNLSST--IIGYRITVV-----AA 1305
QY 665 LNGIPIARPLFFSPDDIKTYGISQFLLGKGVWSPVLKPGV---VSVTAYFPRGNWFD 721
DB 1306 GEGIPFEDFVYS---SVGYTYVTG-----LEPGIDYDISVITLINGE--- 1346
QY 722 LFDYTRSVTASTG-----RYVTLS-----APPDHINVHIQEGNILAMQGKAMTTQ 766
DB 1347 --SAPTTLTQOTAVPPPTDLRFTNIGPDTMRVTWAPPPSIDL-----TNFLVRYSPVKNEE 1400
QY 767 AARK---TPHLLVWMSDCGASFGELFLDDGVSTMGVN-----RGKWTFFVKFI 812
DB 1401 DVAELSISSPDNAVLTN-----LLPGTFYVSVSVSEVFEQHESTPLRGHQK----- 1446
QY 813 AASAKQTCIITSDVVSGEFAVSKQWIDKVTIILGLRKGTKINGYTVRTGAVTRKGDKSKL 872
DB 1447 TGLDSPGTGIDFSDITANSFTV--HWIAPRATITGYRIRHHPEHFSGRPREDRVPHSRNSI 1504
QY 873 KSTPDRKGEFIVAEISGLN-----LILGRE 897
DB 1505 TLTNLTPTGYWVWSIVALNGRESPLLIQOO 1535
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

CM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 ; Search time 48.7158 Seconds  
(without alignments)  
2942.169 Million cell updates/sec

Title: US-10-043-418-3

Perfect score: 4768

Sequence: 1 MKKKIPSLALGILLVFLQY.....VAEISGLNLLGREFKVLVH 903

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*

```

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2884	60.5	901	23	AAU97731
2	2884	60.5	901	23	ABO09151
3	2847	59.7	901	23	AAU97732
4	2847	59.7	901	23	ABO09152
5	2422.5	50.8	877	19	AAW59040
6	2297.5	48.2	682	21	AAV51670
7	1499	31.4	953	23	ABW57174
8	1427.5	29.9	963	20	AAV49895
9	1410.5	29.6	985	18	AAW15191

10	1401.5	29.4	985	14	AAW42395	Glycosyltransferas
11	1401.5	29.4	985	14	AAW42214	Aspergillus niger
12	1378.5	28.9	922	20	AAW88044	An enzyme with sug
13	1375.5	28.8	958	11	AAW07575	Glucosylase encod
14	1374	28.8	958	9	AAW81181	Sequence of glucos
15	1318.5	27.7	1827	20	AAW74090	Human HSI protein
16	1318.5	27.7	1827	22	AAU09028	Human sucrase isom
17	1318.5	27.7	1829	22	ABG14394	Novel human diago
18	1173.5	24.6	1070	13	AAK21521	Alpha galactosidas
19	851	17.8	914	23	ABG69612	Human NOV11c prote
20	848	17.8	914	23	ABG69611	Human NOV11b prote
21	847	17.8	912	23	ABG69613	Human NOV11d prote
22	847	17.8	914	23	ABP52437	Human carbohydrate
23	841	17.6	912	24	AAW79779	Carbohydrate-assoc
24	831	17.4	914	23	AAW66094	Soybean alpha-gluc
25	811.5	17.0	919	18	AAW18580	Potato alpha-glucos
26	811.5	17.0	919	23	AAW66099	S. tuberosum alpha
27	802.5	16.8	941	22	AAW80202	Human protein SEQ
28	802.5	16.8	944	22	AAW79218	Human protein SEQ
29	800.5	16.8	763	23	ABW50119	Listeria monocytog
30	797.5	16.7	967	23	ABG69610	Human NOV1a prote
31	715.5	15.0	565	21	ABW58849	Breast and ovarian
32	708.5	14.9	924	22	ABB67421	Drosophila melanog
33	708.5	14.9	924	22	ABB69429	Drosophila melanog
34	681.5	14.3	718	24	ABP96615	Plasmid pNOV4831 m
35	681.5	14.3	718	24	ABP96625	MAIA fusion protei
36	676	14.2	693	24	ABP96604	S. solfataricus ma
37	676	14.2	712	24	ABP96616	Plasmid pNOV4839 m
38	645.5	13.5	444	21	ABW42826	Human ORFX ORF2590
39	590.5	12.4	235	22	AAW00087	Alpha-glucosidase
40	587	12.3	834	11	AAW04869	Alpha-1-6-glucanas
41	536	11.2	390	23	ABP51380	Human MDDT SEQ ID
42	502.5	10.5	331	22	AAU23490	Novel human enzyme
43	490	10.3	359	22	AAU19420	Human diagnostic a
44	487	10.2	1070	16	AAW77083	M.vulgaris alpha-1
45	487	10.2	1070	16	AAW72712	Alpha-1,4-glucan 1

#### ALIGNMENTS

RESULT 1  
AAU97731  
ID AAU97731 standard; Protein; 901 AA.  
XX  
AC AAU97731;  
XX  
DT 23-AUG-2002 (first entry)  
XX  
DE Common buckwheat alpha-glucosidase polypeptide #1.  
XX  
DE Common buckwheat; alpha-glucosidase; enzyme;  
KW Fagopyrum esculentum Moench.  
XX  
OS Fagopyrum esculentum.  
XX  
PN JP2002065273-A.  
XX  
PD 05-MAR-2002.  
XX  
PF 31-AUG-2000; 2000JP-0262104.  
XX  
PR 31-AUG-2000; 2000JP-0262104.  
XX

(NISO) NIPPON SHOKUHIN KAKO KK.

WPI: 2002-474198/51.

N-PSDB; ABK86269.

A recombinant vector containing alpha-glucosidase gene, and a transformant useful for preparation of alpha-glucosidase -

Claim 2; Page 7-8; 18pp; Japanese.

XX The invention relates to a recombinant vector containing  
CC alpha-glucosidase gene derived from buckwheat. A transformant containing  
CC the recombinant vector can be used for preparing alpha-glucosidase by  
CC culturing the transformant and isolating alpha-glucosidase from the  
CC culture. This sequence represents a buckwheat alpha-glucosidase protein.  
XX  
SQ Sequence 901 AA;  
  
Query Match 60.5%; Score 2884; DB 23; Length 901;  
Best Local Similarity 61.5%; Pred. No. 4.5e-237;  
Matches 559; Conservative 123; Mismatches 183; Indels 44; Gaps 12;  
  
QY 8 LALGILLVFLQYLAVAGISTSENDPEGVIGYGVKSVKVDSTRLTALPOLVKNSSV 67  
Db 14 LLLAATLLFCSLFW-----SESGEVGYGVYVRAKVDS--NTLTAFLKLINASSL 65  
  
QY 68 YGPDIOQLLSITASLESNDRLRVITDAKRRWEIPDNILHR-----HQPPEPPHSLSS 121  
Db 66 YGQDIPNLTFATPEKDYRLRIRITDAEKPRWEIPNEVLRDGSSSHGHQP----- 115  
  
QY 122 LYRTLSSPTNRKRKILLSHPNSDLTFSLINTTPTGFTISRKSTHDLVDA-TPDPTNP 180  
Db 116 ----LDSRPTTPPSAAVLTHPNSDLIFRLHDTNPFSGSVTRRSTNDVLFTRSDPTDP 171  
  
QY 181 TFLIFDQYLHLTSSLPCTRAHYIYGLGEHKSPTQLAHNQTLMRAADIFSSNPVNLVY 240  
Db 172 VGLVFKDQYIQLSSSLFGRRRAHLYGIEGHTKPTFLAHNQTLLWNADIASVNDLNYG 231  
  
QY 241 SHPFYMDVRSSPVAGSTHGVLLNSNGMDVEYTGNRITYKVGIGIIDLFPAGSPQGV 300  
Db 232 SHPFYLDVRAP--LGTSGVLLNSNGMDVEYTGDRITYKVGIGIIDLFPAGTPDEVV 289  
  
QY 301 EQTRVIGRPAPMPYAFQOCQYGHYDVLEQSVAGYAKAKIPLVWMTDIDYMDAY 360  
Db 290 QQYTELIGRPAPMPYAFQOCQYGHYDVLEQSVAGYAKAKIPLVWMTDIDYMDAY 349  
  
QY 361 KQFTLDPNVPEPLDKMKFVNHLKNGQYVILDPGISTN-KTYETIYIRGMKHDVFLKN 419  
Db 350 KQFTLDPNVPEPLDKMKFVNHLKNGQYVILDPGISTN-KTYETIYIRGMKHDVFLKN 409  
  
QY 420 GKPYLGSWPGVYFPDFLKPXSALTFTWDEIKRFLNLLPVDGLWIDMNEISNFISSPPI 479  
Db 410 GEPYQGVWPGVYFPDFLKPXSALTFTWDEIKRFLNLLPVDGLWIDMNEISNFISSPPI 469  
  
QY 480 GSTLDNPPYKINNSGVMLPIINTKIPTTAMHYGDIPEYVNHMLPGYLEARVTRAALIKLT 539  
Db 470 DSPLDNPPYKINNSGVMLPIINTKIPTTAMHYGDIPEYVNHMLPGYLEARVTRAALIKLT 529  
  
QY 540 EKRPFFVLSRSTFGSGKYTAHTWGDNAATWDLVYISIPSMDFGLFGIPMVGADICGFLG 599  
Db 530 KQRPFFVLSRSTFGSGKYTAHTWGDNAATWDLVYISIPSMDFGLFGIPMVGADICGFLG 589  
  
QY 600 NTEELCRRWIQLGAFYPPSRDSSSLGTTQYELRYRWSVAASARKVGLGURYTLPLPYFTL 659  
Db 590 TTEELCRRWIQLGAFYPPSRDSSSLGTTQYELRYRWSVAASARKVGLGURYTLPLPYFTL 649  
  
QY 660 MYEQLNGIPIARPLPESFPDDIKTYGISSQFLLGKGVWSPVLKPGVSVTVAYFPRGNW 719  
Db 650 MYEQLNGIPIARPLPESFPDDIKTYGISSQFLLGKGVWSPVLKPGVSVTVAYFPRGNW 709  
  
QY 720 FDLFDYTRSVTAST-GRYVTLTSAAPPDHNHVIQEGNITLMOGKAMTTOARKTTPFHLVV 778  
Db 710 FNLFDYTKTVSSPTNGSFVTLDALEEINVHVREGSILALQGEAMTTRARNTPFELVV 769  
  
QY 779 MSOC--GASGFEPLDQGVETWGVNRKWTFTVKFIAA-SAKOTCIITSDVSGEFAVSO 835  
Db 770 ISDSGNGSSIGSVFDMGVNDIEMGDGGRWSLVTFSAGLIGNKVTITSSVNGRPAISO 829  
  
QY 836 KWDIKVTILGLRGTGKINGTVV----RTGAVTRKDGKSLKSTPDRKGEFFIVAEISGLN 891  
Db 830 GWKLSKVTILGLSRGSOVKGYTMLSIGRVVTTSVKGARKCSRGT----GKDFVVEIPNIS 885

QY 892 LLLGREFKL 900  
Db 886 LLVGRNFKL 894  
  
RESULT 2  
ABB09151  
ID ABB09151 standard; Protein; 901 AA.  
XX  
AC ABB09151;  
XX  
DT 01-JUL-2002 (first entry)  
XX  
DE Buckwheat alpha-glucosidase protein SEQ ID NO:1.  
XX  
KW Buckwheat; alpha-glucosidase; enzyme; Fagopyrum esculentum Moeench;  
XX  
OS Fagopyrum esculentum.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..31  
FT Protein /label= signal  
FT 32..901  
FT /label= alpha\_glucosidase  
XX  
PN JP2002065272-A.  
XX  
PD 05-MAR-2002.  
XX  
PF 31-AUG-2000; 2000JP-0262102.  
XX  
PR 31-AUG-2000; 2000JP-0262102.  
XX  
PA (NTSO ) NIPPON SHOKUHN KAKO KK.  
XX  
DR WPI; 2002-346763/38.  
DR N-PSDB; ABL51446.  
XX  
PT A buckwheat-derived alpha-glucosidase gene -  
XX  
PS Claim 1; Page 7-9; 21pp; Japanese.  
XX  
CC The present sequence represents an alpha-glucosidase isolated from  
CC buckwheat (Fagopyrum esculentum Moeench). The alpha-glucosidase gene  
CC can be used for the preparation of alpha-glucosidase derived from  
CC buckwheat.  
XX  
SQ Sequence 901 AA;  
  
Query Match 60.5%; Score 2884; DB 23; Length 901;  
Best Local Similarity 61.5%; Pred. No. 4.5e-237;  
Matches 559; Conservative 123; Mismatches 183; Indels 44; Gaps 12;  
  
QY 8 LALGILLVFLQYLAVAGISTSENDPEGVIGYGVKSVKVDSTRLTALPOLVKNSSV 67  
Db 14 LLLAATLLFCSLFW-----SESGEVGYGVYVRAKVDS--NTLTAFLKLINASSL 65  
  
QY 68 YGPDIOQLLSITASLESNDRLRVITDAKRRWEIPDNILHR-----HQPPEPPHSLSS 121  
Db 66 YGQDIPNLTFATPEKDYRLRIRITDAEKPRWEIPNEVLRDGSSSHGHQP----- 115  
  
QY 122 LYRTLSSPTNRKRKILLSHPNSDLTFSLINTTPTGFTISRKSTHDLVDA-TPDPTNP 180  
Db 116 ----LDSRPTTPPSAAVLTHPNSDLIFRLHDTNPFSGSVTRRSTNDVLFTRSDPTDP 171  
  
QY 181 TFLIFDQYLHLTSSLPCTRAHYIYGLGEHKSPTQLAHNQTLMRAADIFSSNPVNLVY 240  
Db 172 VGLVFKDQYIQLSSSLFGRRRAHLYGIEGHTKPTFLAHNQTLLWNADIASVNDLNYG 231  
  
QY 241 SHPFYMDVRSSPVAGSTHGVLLNSNGMDVEYTGNRITYKVGIGIIDLFPAGSPQGV 300  
Db 232 SHPFYLDVRAP--LGTSGVLLNSNGMDVEYTGDRITYKVGIGIIDLFPAGTPDEVV 289









Db 507 HQQVDFGMMMDNERNFNRVRSQQGCPNNELENNPY---XPGVVGIIQLAAATICASSHQ 563  
 QY 512 GDPEYNVNLFCYLAARVTRALIKLTKRPVLRSSTFGSGKYTAHTWTDGNAATWMD 571  
 Db 564 FLSTHNLNLYGLTEAIASSRALVKTGRTPVRSRSTFGHGRVAGHMTEDVRSWEH 623  
 QY 572 LVYSIPSMIDFGLGIPMVGADICGFLGNTTELCRRMIQLGAFYFSDHSLGTYQE 631  
 Db 624 LAYSVDILOFNLLGVPLVADICGFIGDTSELCVRWTLQGAFFPMENHNDLSVQE 683  
 QY 632 LYRWESVAASA-RKVLGLRYTLPPYTLMYEQLNGIPIARPLPFSDDIKTYGISQ 690  
 Db 684 PYRFSETAQAMKAFALRYALLPYLTIFHRAHVRGDTVARPLFLEFREDPSTWSVDQ 743  
 QY 691 FLGKGVMSPVLKPGVSVTAVERPGRMFDL-----FDYTRSVTASTG 734  
 Db 744 LMGVALLITPVLEPKTEVTGFGKGTWYNMQUVSVDLGLTLPSSASSFRSAVQSG 803  
 QY 735 RYVTLAPPDHHNVHIOEGNILAMOGKAMTQAAKTPPHLLVMSDCGASFGELFLDDG 794  
 Db 804 QWLTLEAPLDTINHLREGVILPLOGPSLTSTTESRKQPMALAVALTASGEADGELFWDDG 863  
 QY 795 VEVTMGVNRKWTFFVKFIAASAKOTCIITSDV-VSGEFVAVSQKWIDKVTILGLRKGT 853  
 Db 864 ESLAV-LEHGAYTLVTF---SAKNTIVNKLVRVTKEGAELO---LKEVTVLGVA----- 911  
 QY 854 NGYTVRTGAVTRKGDGSKLSTPKRGEFIVAEISGLNLLGREFKL 900  
 Db 912 ---TAPQVLSNG:PVSNTFYSNDKSLAI-----PVSULMGELOI 950

## RESULT 8

AAY49895  
 ID AAY49895 standard; Protein; 963 AA.

AC AAY49895;

XX 28-JAN-2000 (first entry)

XX Endomyces fibuliger glucoamylase GLA2 protein.

XX Endomyces fibuliger; glucoamylase; GLA2; hydrolysis; carbohydrate;  
 KW glucose; alcohol; starch; leguminous plant; cooked; oleaginous plant;  
 KW roasted; cereal; soya; wheat; rice.

XX Saccharomycopsis fibuligera.

XX EP959130-A1.

XX 24-NOV-1999.

XX 06-MAY-1998; 98EP-0201473.

XX 06-MAY-1998; 98EP-0201473.

XX (NEST ) SOC PROD NESTLE SA.

XX Pridmore RD, Kochhar S;

XX WPI; 1999-622099/54.

XX N-PSDB; AA232388.

XX New recombinant glucoamylase from Endomyces fibuliger, useful for  
 PT hydrolyzing carbohydrate-containing materials -

XX Claim 1; Page 13-15; 20pp; English.

XX The present sequence represents Endomyces fibuliger (also called

CC Saccharomycopsis fibuliger) glucoamylase GLA2. The invention describes  
 CC the recombinant GLA2 enzyme, which is useful for hydrolysing  
 CC carbohydrate-containing materials, especially materials containing rice  
 CC carbohydrates. The glucoamylase is used to produce glucose and alcohol  
 CC from starch containing raw materials. GLA2 is useful for hydrolysing

CC carbohydrate containing materials e.g. mixtures of a source of proteins  
 CC and a source of carbohydrates, especially a mixture of leguminous plant  
 CC or of a cooked oleaginous plant and of a cooked or roasted cereal source  
 CC e.g. a mixture of soya or cooked beans and of cooked or roasted wheat or  
 CC rice.

XX Sequence 963 AA;

QY Query Match 29.9%; Score 1427.5; DB 20; Length 963;  
 Best Local Similarity 34.7%; Pred. No. 1.8e-112;  
 Matches 343; Conservative 165; Mismatches 310; Indels 171; Gaps 31;

QY 8 LALGILLVFLQLVLAGISTSENDEPGVIGYKVKSVKVDSCGTRSLTALPOLVKNSSV 67  
 Db 48 LTLGVSV---PHVVKATADANE---AAKGVELVNV---TTTAKGLTGILKLNENAI 97  
 QY 68 YGDIQLLSTASLENDRLRVKITDAKRRWEIPDNILHRHOPPPPPHSLSLRYTL 127  
 Db 98 YGYDFDYLNLSVEYQSDRLNVHI-----EPVTDNVFILPESLV 137  
 QY 128 SSPIT-NRRKILLSH--PNSDLTFSLINITPFGFTISRKSTHVDVLFDPDTPNPTFL 184  
 Db 138 AKPSADGDKIESFHFGGSDLVFE-YSSKNQFGEILRKSTGKSIESTGNP-----LV 190  
 QY 185 FIDQYLHLTSSLPTTRAH-IYGLGESHKPTFQLAHNQTLTMRADIPSSNP-DVNLVGS 242  
 Db 191 FSNQIQFNISLP--KDHFTITGLGE-SIHCFRNEPGIVKTLYANDI--ANPIDGNIYGVH 245  
 QY 243 PFYMDVRSSPVAGSTHGVLLNNGMDVEYTNRIYKVIIGIIDLFFAGSPGQVVEQ 302  
 Db 246 PFYIDQRFD--TNATHGVYWRVTSIOEAVAGNESLWRLSGVVDLYFFSGPKPKDVIQ 303  
 QY 303 FTRVIGRPAPMPYMAFGQOCRYGHDVVELQSVAGYAKAKIPLFVMTDIDYMDAYKD 362  
 Db 304 YVKEVGLTFPTQVWALGYHOCRWGYDTIELDRVENFKKFDIPJTIWSDIDYMSYKD 363  
 QY 363 FTLDVNVFLDKMKXVFNHLHNGQKYVWILDPGI-----STNKTYETIRGMKHDFV 415  
 Db 364 FTNDPHRYPLEKYQQFLDKLHNNQHYVPIIDAAIVVPENPNATDNDYDVFYGNETDVF 423  
 QY 416 LKR-NGKPYLGSVMPGVPPDFLKPALSALTFTWDEIKRFLNLLPVDGLWIDMNEISNF-- 472  
 Db 424 LKMPDGLSYIGAVMPGYTFVFPDFLSENIQKYMTKVKFQWYQQIKFGIWLDMNEVSSFCV 483  
 QY 473 -----ISSPPI----- 478  
 Db 484 GSCSGKITDNVPHPPFAVCGEATEPEGPNKTNCTEYASFSSLAASPTSDSDSVSS 543  
 QY 479 -----FG-STLONPPYKINNSGVMLPIINKTIPPTAMHYGDIPEYNVHMLFGY 525  
 Db 544 TSASIDSLNTLAPKGNINYPYAINNDQGDHDLATHAVSPNATHQOGLTEYDVHNLGY 603  
 QY 526 LEARVTRAAALIKL-TEKRPFLSRSTFGSGKYTAHTWTDGNAATWDLVYSIPSMIDFGL 584  
 Db 604 LETNATFEALLEIQPNKRPFIISRSFAGSGRGTGHWGDNYSQFRSAIFSIAQAFSGL 663  
 QY 585 FGIPWVGADICGFLGNTTELCRRMIQLGAFYFSDHSLGTYQELYRWESVAASARK 644  
 Db 664 SGIPFFGADVCGFNGSDYELCSRWMLGSFPFYRNHILGALISQEPYVWESVTEATKT 723  
 QY 645 VLGLRYTLLEPYFTLMEYEAQLNGIPIARPLPFSDDIKTYGISSQFLLGKGVMSVPLK 704  
 Db 724 SMOIRYLLPYTYTLLEHAHTTGIPILRAPAWQPKENKNVSTVDTOFFVGDALVVTVPLE 783  
 QY 705 PGVSVTVAYFP-RGN---WFDLFYTRSVTASTGRVYVTLAPPDHHNVHIOEINLAWQ 760  
 Db 784 QGVDTVKGTFPGSGNEEVYDWTYTBKQ-NFTDGKNETLQAPLGHIPJHIRGHILPTOE 842  
 QY 761 KAMTQAAKTPPHLLVMSDCGASFGELFLDDG-----VEVTMGVNRKWTFFVKFIAASA 816  
 Db 843 PAVTTTESRQNPGLIVALDKCKAGSKLYSDDGSEVEVESL-----FWNFASD- 893  
 QY 817 KQTCITSDVVSVEFAVSQKWIDKVTILGLRKGTINKINGYTVRTGATVRKGDKSLKSTP 876



```

251 QY SPVAGS-----THGVLLNSNGMOVETVGNKRTVVKVYGGIIDL 289
249 Db YKVGQNKSHTVKSSSEAPSOEYVSYSVGHGFTFRNAGHQEILLRDQKLIWRTLGGSDVLT 308
290 QY FPAQSPGOVVEQF-TRVIGRPAMPYWAEGFOCRVGYHVDVYELQSVVAGYAKAPILE 348
309 Db FYSGTQAEVTKQVLSTVGLUPAMQQVNTLGFQCRWGVNNWSEFEDVLANFERFPILE 368
349 QY VNMUTDIDYMDAYKDTLDPVNFPLDKMKXVNNLHNKQKQYVVILDPGI-----STNKT 402
369 Db YLMADIDYMGYRNFNDQHEFSYECEKFLNKLHAGGRWVPVLDGALYIPNPENASDA 428
403 QY YETVIRGMKHDFVLR-NGKPYLGSVMPGPVYPDFLKPSSALFTWDEIKRFLNLLPVDG 461
429 Db YETYDRGAKDVFVKNPDGSLYIGAVWPGYVTPDWHHPKASDFWANELVTWNKLYHDG 488
462 QY LWMIDWNEISNF-----472
489 Db VMDYMAEYSSFCVSCGCTGNLSMNPANPPALPGEQGNVYVYDPEGFTNTEAASASA 548
473 QY -----ISSPPIPG-STLDNPPYKINNSGVMLPIINKTTPPTAMHYGD 513
549 Db GAASQSAASSTTTTSAPYLRTPTTPGVNRVDHPYVINVQPGHDLSVHAISPNSTHSDG 608
514 QY IPEYVNRMLPGYLBARVTRAAALIKLTE-KRPFVLSRSTPSSGSKYTAHTGTDNNAATWDL 572
609 Db VOBYDVHSLYHGQINATYHGLLKWENKRPFIARSTFSGSKGAWHGWDGNSXWGM 668
573 QY VYSIPSMLDGFLGFI PMVGADICGFLGNTTBELCRMWIQLGAPYPSRDHSSLGTTYQEL 632
669 Db FFSISQALQFSLGFI PMFGVDTCFGNTDBELCNRMWLSAFPFYRHNVLUSALPQSP 728
633 QY YWESVAASARKVLGLRYLLPYEYTYLMBYLAQLNGPIARPLFFSFPDDIKTYGSSQFL 692
729 Db YRWASVIDATKAAMNIRYAILPYEYTYFLHLAHTTGSTVMALAWEPNDPSLAAVGTQFL 788
693 QY LQKGMVSPVLKPGVWSVTAFF---RGN-WFLDFYTRSVTASTGRYVTLSPAPPHINV 748
789 Db VGPSVMVIVPLEPQVDTVQGVFPVGHGEVMDYWSQT-AVDKAPGVNTTISAPLGHIPV 847
749 QY HJQEGNILLAMQCKAMTTCQARKPTPHLLVVMSCGASGELFLDDGVEV---TMCVNRG 804
848 Db FVRGSIILPMQOEVALTTTRDKARTPWSLLASUSNGTASGQYLDGGEVVPEDTLSDV--- 905
805 QY KMTFVKFIA-----ASAKQTCIITSDVWSGEFAVSQKW---VIDKVTILGLRK---GT 851
906 Db -----FLASRSTLRASARGT-----WKEANPLANVTIVGVTEKPSSV 942
852 QY KINGTVTRTGA 863
943 Db TLNGETLSSDSV 954

```

RESULT 10	
AA342995	
ID	AA342995 standard; Protein; 985 AA.
XX	
AC	AA342995;
XX	
DT	16-MAY-1994 (first entry)
XX	
XX	Glycosyltransferase.
XX	
KW	Glycosyltransferase; GT; Grase; Aspergillus niger; plasmid;
KW	transformation; expression.
XX	
OS	Aspergillus niger.
XX	
PN	JP05268951-A.
XX	
PD	19-OCT-1993.
XX	



Db 47 LIANI-----DDFOADAQSCPCYKASKQVHNS---RGFTASQLAGRCNVYGTVESL 99  
 Qy 76 SITASLESNDRLRVRI-----TDKARHWE-IPNIIHRHQPPPPHSLSSLYRTLSSP 130  
 Db 100 TLSVEYQDSRLNIQIPLTHVDSTNASWFLSRLV-----PRPKASLNA----- 144  
 Qy 131 TTNRRKILLSHPNSDLTFTSLINTTPGFTISRKSTHDLVDFATPDPTNPNTLFIIDVYL 190  
 Db 145 -----SVSOSDLFVSWNSNPFNFVKIRKATGDAIF-----STEGTLVLYENQFI 189  
 Qy 191 HLTSSLPGTRAHIYGLGESHKPTFOLAHNQTLTMRADIPSSNP-DVNLYGSHPFYMD-- 247  
 Db 190 EFTALP-BEYNLYGLGEHI-TQRLQRANLTIYPSD--DGTPIQNLYGQHPFLDTR 245  
 Qy 248 -----VRSPVAGS-----THGVLLNSNGMDVEYTNRIYKIVGIGIDL 288  
 Db 246 YYKGRQNGSYIYFKVSEADASQDYISLHGVLFRNSHGLEILLRSQKILWRTLGGIDL 305  
 Qy 289 YFPAGSPGOVWEQF-TRVIGRAPMPYMAFGPOQCRYGVDVYELQSVVAGYAKAKIPL 347  
 Db 306 TFSYGPAPADVTRQYLTSTVGLPAMQOYNTLGFHQCRWGNWNSDLADVVANFEKEIPL 365  
 Qy 348 EVNMTDIDYMDAYKDFLTPDVPNPLDKMKKPVANLHKNGQYVILDPGI-----STNK 401  
 Db 366 EYIWTIDIMHGYNFNDNDRFSYSEGDFLSKLHESGRYVPIVDAALYIPNPENASD 425  
 Qy 402 TYETIIRGMKHDFLKR-NCKPYLGSWPGVPYFPDLKPSALTFTWDEIKFLNLLPYD 460  
 Db 426 AYATYDRGAADDVFLKNPDGSLYGAVMPGYTFPDWHHPKAVDFWANELVMSKKAEPD 485  
 Qy 461 GLWIDMNEISNF-----ISSPPIPG-STLDNPPYKINNSGVMLPIINKTIPPTAMHY 511  
 Db 486 GVWYDMSEVSFCVSGCGTGNLTNPAHPSFLPGBPGDIIYDYPAPNITWATEASAS 545  
 Qy 473 -----ISSPPIPG-STLDNPPYKINNSGVMLPIINKTIPPTAMHY 511  
 Db 546 AGASSQAAATATTSTSVSYLRTPTPGVRNVHPYVINVHQEGHDLVSHAVSPNATHV 605  
 Qy 512 GDPIPEYNVNLFGYLEARVTRALIKL--TEKPPVLSRSTFGSGKYTAHTNGONATW 569  
 Db 606 DGVEEDVHGLYHQGLNATYQGLLEVMHSHKRPFIIGRSTFAGSGKAGHGGGNYSKW 665  
 Qy 570 NDLVYIPSMDFGLGIPMGVADICGFLGNTTEELCRRWIQLGAFYFSPSRDHSSLGTTY 629  
 Db 666 WSMYYSISQALSFLGIPMEGADTCFNGNSDEELCNRMQLSARFPFYRNHNLSTIP 725  
 Qy 630 QELYRWESVAASARKVLGRLYTLPIFYTLMYEAQUNGPIAPPLFPSPDDIKTYGISS 689  
 Db 726 QEPYRWASVIEATKSAMRIYAILPYEYTLFDLAHTTGSTVVRALSWEPFPNDPTLAAVET 785  
 Qy 690 QFLGKGVMVSPVLKPGVSVTAPEP---RGN-WFDLFYTR-SVTASTGRYVYTLAPPD 744  
 Db 786 QFMVGPAIMVVPVLEPLVNTVKVFGVGHGEVWYDW--YTQAQVDAKPGVNTTISAPLG 843  
 Qy 745 HINVHIQEGNILAMQKAMTTOARKTTPHLLVMSDCGASGEFLDDGVEV---TMG 800  
 Db 844 HIPVYVRGNNLPMQEPALTTREARQTPWALLAALGNSGTASQLYLDGESIYPNATLH 903  
 Qy 801 VN-----RKWTFVKFIAASAKQTCIITSVDSVSGEFAVSQKWIVDKVITLGR 848  
 Db 904 VDFATSRSSLSRQAQGRW-----KERNPLANVTVLGVN 936  
 Qy 849 K---GTKINGYTVRTGAVT 864  
 Db 937 KEPSAVTLNGQAVPGSVT 955

RESULT 12

AAW88044

ID AAW88044 standard; Protein; 922 AA.

XX

AC AAW88044;

XX 09-APR-1999 (first entry)  
 DT An enzyme with sugar transferase activity.  
 DE Acetomoniom sp. S4G13; glucose transfer; sugar transferase  
 KW sugar receptor; starch.  
 XX Acetomoniom sp.  
 OS JP11009276-A.  
 PN 19-JAN-1999.  
 XX 19-JUN-1997; 97JP-01631110.  
 PF 19-JUN-1997; 97JP-01631110.  
 PR (KIRI ) KIRIN BREWERY KK.  
 PA MPI; 1999-145893/13.  
 DR N-PSDB; AAX04639.  
 XX New sugar transferase gene and enzyme - useful for catalysing the  
 PT transfer of an alpha-1 right arrow 3 bond to a sugar receptor for  
 PT saccharide preparation  
 XX Claim 1; Pages 13-17; 20pp; Japanese.  
 PS The present sequence represents a sugar transferase protein of Acetomoniom  
 XX sp. S4G13. The protein preferably catalyses the glucose transfer of an  
 CC alpha-1 right arrow 3 bond or the glucose transfer of an alpha-1 right  
 CC arrow 3 and an alpha-1 right arrow 4 bond to a sugar receptor by  
 CC reacting with a substrate selected from starch and its decomposition  
 CC products.  
 XX Sequence 922 AA;  
 SQ  
 Query Match 28 9t; Score 1378.5; DB 20; Length 922;  
 Best Local Similarity 34.3t; Pred. No. 2.6e-108;  
 Matches 324; Conservative 140; Mismatches 311; Indels 169; Gaps 26;  
 Qy 39 GYKVKSVK-DSGTRRSITLALPOLVKNSSVYGPDIQLLSITASLESNDRLRVITDAKHR 97  
 Db 36 GYAASNQVSDTGLTADTLAGE---PCDAYGEDLKDLLEVYETETENLHVKIQDKGNQ 92  
 Qy 98 RWEIPDNILHRHQPPPPHSLSSLYRTLSSPTNRRKILLSHPNSDLTFTSLINTTFFG 157  
 Db 93 VYQIPESVF-----PRPGSIDP-----ESSIRFAYAE-BPPS 125  
 Qy 158 FTISRKSTHDLVDFATPDPTNPNTLFIIDQYLHLTSSLGCTRAHIYGLGESHKPTQLA 217  
 Db 126 FNITRADTDEVLFTSA-----ASIVFESQYLRLTSIP-TDPYLYGLGARNDPMLES 178  
 Qy 218 HNQTLTMRADIPSSNPDVNLVYGHSHPFYMDVRSSPVAGSTHGVLNLSNGMDV-----EY 272  
 Db 179 VGYIRTFWNQDSYGVPGNANLYGSHPVYIDHREI---GTHGVLFNLSNGMDVLI DEDEE 234  
 Qy 273 TGNRITYKVIIGIIDLFFPAGSPGQVVEQFTRVIGRPAPMPYMAFGQCRYGYHDVYE 332  
 Db 235 GGYLEYNTLGLVDLYFFVFGDSPKAVEEYETAGRPMPQMPYMGGLFHOCKYGYQDAFM 294  
 Qy 333 LQSVVAGYAKAKIPLVYMTDIDYMDAYKDFLTPDVPNFPDLKMKKFNYNLHKNGKTVVI 392  
 Db 295 VAEVYVYNSQABIELEVYMTDIDYMDRRRRVFTVDPRFPLKIRAVVDYLHEHDQRVYVM 354  
 Qy 393 LDPGISTNKTYETIIRGMKHDFLKR-NCKPYLGSWMPGPVYFPDFLPSALTFTWDEIK 451  
 Db 355 VDPALAYVES-GTLDRGLDDEDFLLRSNGSVWLGVPWPGVTFVDFDFAENITQYWNNEFA 413  
 Qy 452 RFLNL---LPVDGLWIDMNEISNF-----ISSPP----- 477  
 Db 414 LFFDADEGVDDIGLWIDMNEISNPFPCNDPNYEAAGFPFPFPFPFPFPFPFPFPFP 473

Qy 478 -----IPGSTL 483  
 Db 474 LOPEGTCEGETAGSSKRDGSPGQLVTRQPGFSRPHRPHRQVEYEDGQKLPGRDL 533  
 Qy 484 DNPYPKINNSGMLP-----IINKTPTTAMHYGDIPEYVNNLFGYLEARVTRAL 535  
 Db 534 LYPEYAIHNKAAFRDDWNADKGGISNKTNTVNIHQGLAEYDVHNLGAMSSASRDAM 593  
 Qy 536 -IKLTEKRPVLSRSFSGSGKYTAHMTGDNAATWDLVYSIPSMDF-GLFGIPMGAD 593  
 Db 594 EARRPGLRPFITRSTFPHAGSKVGLWGLDNLNMQYRESIRTMLAYTSIFQGVGSD 653  
 Qy 594 ICGFLGNTTEELCRRIQLGAFYPPFSRDHSSLCTTYQELYRMESVAASARKVGLRYTLL 653  
 Db 654 VCGFGDTEELCARWASLGATQTFERNHAQYEAQVPEFYQWESVAESARRAIGARYLL 713  
 Qy 654 PFTYLMYEAQLNGIPIARPLPSPDDIKTYGSSQFLGKGVVSPVLKGVVSVTAY 713  
 Db 714 DMYTALWKEQSGTPAVVPMFYFPEPDKGTLELQYFYGPGLVAPVVEQGSTVDVY 773  
 Qy 714 FPRGNWFLDFDY--TRSVTASTGRVYVLSAPPDHNINVHIQEGNILAM-OGKAMTTCAARK 770  
 Db 774 LREGKVF--YDWTHERAIOGEGSGSVTGVTMTMPLFIRGGVILPLRENSAMTTLEURK 831  
 Qy 771 TPFHLLVMSDCGASRGELFLDDGVETVMGVNRGKMTFVKFIAASAKQTCIITSDWSGE 830  
 Db 832 EKPELLVALDNDGKAKGELYIDG---ESLQGESYAVKFEYAHG-----VVTLD---GE 880  
 Qy 831 FAVSQKWVDKVTILGLR-KGTYKINGYVTRTGAVTRKGDGSKLK 873  
 Db 88: F--SEDPVPEVASVLLRPKGKEI---VVEVGKSFETAGGRILK 919

## RESULT 13

AA07575  
 ID AA07575 standard; protein; 958 AA.

AC AA07575;

DT 25-MAR-2003 (updated)

DT 02-FEB-1991 (first entry)

DE Glucoamylase encoded by GAM1 gene.

KW Schwanniomyces yeast cells; expression cassette; glucoamylase;  
 KW marker gene; regulon; signal peptide; terminator;  
 KW autonomously replicating sequence.

OS Schwanniomyces occidentalis.

PN BP394538-A.

PD 31-OCT-1990.

PF 28-APR-1989; 89EP-0107780.

PR 28-APR-1989; 89EP-0107780.

PA (RHEI-) RHEIN BIOTECH GES.

PI Hollenberg C, Strasser A;

DR WPI; 1990-328670/44.

DR N-PSDB; AA006389.

XX Transformed Schwanniomyces yeast cells - contg. an expression  
 PT cassette contg. regulon, DNA coding for foreign protein and  
 PT terminator

PS Disclosure; Fig 2B(1-9); 59pp; English.

XX The expression cassette may contain a regulon (R), a signal peptide

CC sequence (S), a foreign protein sequence (F) and a terminator (T)  
 CC of which R and/or S and/or T are derived from the glucoamylase gene  
 CC of S. occidentalis.  
 CC R is pref. a 1.3 kb BamHI-PvuII fragment (bases -1 to -320 of this  
 CC sequence).  
 CC S contains all or part of one of the sequences represented  
 CC in AA006383-87.  
 CC F is e.g. a cellulase, interleukin, insulin-like-growth factor,  
 CC interferon etc.  
 CC T is pref. all or part of the terminator of this sequence  
 CC (bp 2875-3320).  
 CC GAM1 acts as selective marker for the vector carrying the expression  
 CC cassette.  
 CC R, S, F and T may be modified by insertions, deletions or substitutions  
 CC such that biological activity is retained or improved.  
 CC See also AA006388.  
 CC (Updated on 25-MAR-2003 to correct PA field.)  
 XX  
 SQ Sequence 958 AA;

Query Match 28.8%; Score 1375.5; DB 11; Length 958;  
 Best Local Similarity 33.4%; Pred. No. 4.9e-108;  
 Matches 332; Conservative 142; Mismatches 316; Indels 203; Gaps 28;

Qy 3 KKIPSLALGILLFELLOYL-----VAGISTSENDPEGVI----- 36  
 Db 5 KLIKSVIGLGLVSAIQAAAPASSIGSSASASSESSQATIPNDVTLGVKQIPNIFNDSA 64  
 Qy 37 -----GYGVKSVKVDSTRRSLTALPOLVKNSSVYGPDIQLLSITASLEGNRLVR 90  
 Db 65 VDANAAAKGYDLNVN---TNTPRGLTGILKLEATNIYGYDFDYLNLTVQEADTRLNVH 121  
 Qy 91 ITDAKRRRWIIPDNLHRHQPFPHPHSLSSLY---RTLLSSPTNRKILLSHPNSDLT 147  
 Db 122 I-----EPTDLSDFVLPEHLVVKPLVEGDAQSYNDFNSDLV 158  
 Qy 148 FSLINTPFGFTISRKSTHVDLFDATPDPTNPNTFLIFDQYLHLTSSLPGTRAHIYGLG 207  
 Db 159 FEYSN-TDFSEVIRSTKEVLFSTKGNP-----LVFSNQFIQFNSSLPNHV-ITGLG 210  
 Qy 208 EH-----SKPTFQLAHNQTLTMAADIPSSNP-DVNLVYGHSPFYMDVRSFVAGSTHVL 261  
 Db 211 ESIHGLVNEP-----GSVKTLPANDV--GDPIDGNIYGVHPVYLDQRYD--TETHAVY 260  
 Qy 262 LLNSGNDVETVGNRTYKVIIGIIDLFPAGSPGVQVQFTRVIGRPAPMPYWAQFQ 321  
 Db 261 WRTSAIOEVLIGEISTWALSGVIDLYFFSGPTPKDAIQQYVKEIGLPAFAQYWSLGYH 320  
 Qy 322 QCRGYHDVYELQSVAGYAKAKIPLRVMTDIDYMDAYKDFLDPVNPFLDKMKKPVNN 381  
 Db 321 QCRWGYDTIEKLSFVVENFKFNIPLETIMSDIDYMSYKDFYDHRFFLDEYRKELDE 380  
 Qy 382 LHKNGQKVVVILD-----PGISTNKTETYIRGMKHVDVFLKR--NGKPYLGSVWPGEVY 433  
 Db 381 LHKNGHYVPIIDAAIYVNPNNATNEYQPFHYGNETDVFLENKPDGSLYIGAVWQ-VTL 439  
 Qy 434 FPDFL--KPSALTFTWDEIKRPLNLLPVDGLWDMNEISNP-----ISSPPIP-- 479  
 Db 440 FSRFLSRKHSMD---DKVIKDWYELTFDFGIWADMNEVSSFCVSGCGTKGYFENPAYPPF 496  
 Qy 480 -----GSTLD----- 484  
 Db 497 TVGSKATSYVPFDPVSNASEMKSIQSSISATAKTSSVSSSSSTIDYMTLAPCKGNI 556  
 Qy 485 -NPPYKINNSGVMLPIINKTIPPTAMHYGDIPEYVNNLFGYLEARVTRALIKL-TEKR 542  
 Db 557 NYPPYAIYNNQGDSDLATHAVSPNATHADGTVEYDIHNLYGYLQENATYHALLEVFNKR 616  
 Qy 543 PFVLSRSTFSGSGKYTAHWTGDNAATWDLVYSIPSMDFGLFGIPMWGADICGFLGNTT 602  
 Db 617 PFMISRSTFPRAKWTGHGWDNTADWAYAYFIPQAFSMGIAGLFFGADVCGFNGNSD 676  
 Qy 603 EELCRRWIQLGAFYPPFSRDHSSLGTTTYQELYRMESVAASARKVGLRYTLFPYTIMYE 662

Db 677 SELCSRWMLGSGFFPYRNHNYLGAIDQEPYVWESVAEARTSMARIRYLLPPYYTLJHE 736  
Qy 663 AOLNGIPARLPFPSPDDIKTYGISSQFLGKGMVSPVLKPGVSVYATYPRGN---- 718  
Db 737 SHITGLPILRAFSGWQFNDRLSGVONQFFVGGGLVVTPLVPGVDKVKGVFPAGKEEV 796  
Qy 719 WFDLFDYTRSVTASTGRYVTLSPAPDHINVHIOEGNILAMOGKAMTTOARKTPPHLLVV 778  
Db 797 YDWTY-OREVHFQKGNELDAPLGHIPLHIRGGNVLPTEPGYTVAESRQNPFGGLIVA 855  
Qy 779 MSDCGASFGELFLDDGVEVTMGVNRGKMTFVKFIAASAKQTCIITSDVVSGEFAVSQKVV 938  
Db 856 LDNDGKAQGLYLDGGSVLVDSS-----LLVSFVSNDTILSASP--SGDYKADQP-- 904  
Qy 839 IDKVTILGLRKGTKINGYVTRTGAIVTRKGDGSK 871  
Db 905 LANVTILGV-----GDKPK 918

## RESULT 14

AA81181  
ID AAP81181 standard; protein; 958 AA.

XX AAP81181;

XX 25-MAR-2003 (updated)

DT 12-DEC-1990 (first entry)

XX Sequence of glucoamylase encoded by genomic DNA of Schwanniomycetes  
DE castellii.

XX Brewing; beer; breadmaking; biomass.

XX Schwanniomycetes castellii.

XX EP260404-A.

XX 23-MAR-1988.

XX 17-JUL-1987; 87EP-0110370.

XX 17-JUL-1987; 87EP-0110370.

XX (HETB ) HEINEKEN TECH BEHEER NV.

XX Strasser A, Martens FB, Dohmen J, Hollenberg CP;

DR WPI; 1988-078794/12.

DR N-PSDB; AAN81526.

XX Polypeptide prodn. from cells transformed with yeast DNA -  
PT esp. coding alpha amylase or glucoamylase, able to convert  
PT starch to ethanol

XX Claim 19; Fig 17A-17E; 74pp; English.

XX A cosmid library is constructed from genomic DNA of ATCC26076 and this  
CC screened for inserts contg. the AA gene by ability to transform  
CC S.cerevisiae GRF 18 to histidine prototrophy. One positive cosmid  
CC (32kb; pXcl) was digested with EcoRI, religated and used to transform  
CC E.coli JA221. A GA-contg. cosmid (about 40 kb) was isolated. This was  
CC cut with BamHI and religated to give the plasmid pCUD5-AMG1 contg. a  
CC 12kb fragment (in AAN81526).  
CC (Updated on 25-MAR-2003 to correct PA field.)

XX Sequence 958 AA;

Query Match 28.8%; Score 1374; DB 9; Length 958;

Best Local Similarity 33.8%; Pred. No. 6.6e-108;

Matches 328; Conservative 142; Mismatches 315; Indels 184; Gaps 27;

Qy 3 KKTPLSALGILLVFLQYL-----VAGISTENDEPGVI----- 36

Db 5' KLIKSIIVIGLGLVSAIQAPASSICGSSASASSSSSQATIPNDVTLVGKQIPNIFNDSA 64  
Qy 37 -----GYGKVKSVKVDSGTRSLTALPOLVKNSSVGYPDQILLSITASLESNDRLVR 90  
Db 65 VDANAAGYDVLNV---TNTPRGLTGILKKEATNIYGYDFDYLNLTVQEADRLNVH 121  
Qy 91 ITDAKRRRWEIPDNILHRHQPPPPPHSLSLY---RTLSSPTTNRKRLKILLSHNSDLT 147  
Db 122 I-----EPTDLSDFVLPEHLVVKPLVEGDAQSYNFNDSLV 158  
Qy 148 FSLINTFPFGFTISRKSTHDVLPDPTNPNTFLIFIDQYLHLTSSLPCTRAHIYGLG 207  
Db 159 FEYSN-TDFSEVIRSRSTKEVLFSTKGNP-----LVFSNQFIOFNSSLPNHV-ITGLG 210  
Qy 208 EH-----SKPTFOLAHNQTLTMRADIPSNP-DVNLVYGSHPFMDVRSVPVAGSTHVL 261  
Db 211 ESIHGLVNEP-----GSVKTLFANDV--GDPIDGNIYGVHPVYLDQYD--TETTHAVY 260  
Qy 262 LLNSGMDVEYTGRIITYKYVIGGIIIDLYFFAGSPGVQVVEQFTRVIGRPAPMPYWARFGQ 321  
Db 261 WRTSAIOEVLIGEESITWRALSGVIDLYFFSGPTPKDAIQOYVKBIGLPAPQYMSLYGH 320  
Qy 322 QCRYGHDVYELQSWAGYAKAKIPIEVMWMTDIDYMDAYKDFLDVNFPLDKMKKFWNN 381  
Db 321 QCRWGYDTIEKLSVENVFKFNPILETIWSDDIDYMSYKDFYDHRFPLDEYKFLDE 380  
Qy 382 LHKNGQKTVILD-----PGISTNKTYYTIRGMKHDVFLKR--NGKPYLGSVWPGFVY 433  
Db 381 LHKNHQYVPLDAAIYVNPNNATNEYQPFYHNETDVLKNPDGSLYIGVMQ-VTL 439  
Qy 434 FPDFL--KPSALTEWTDIEIKFELNLLPVDGLWIDMNEISNF-----ISSPPIP-- 479  
Db 440 PSRFLSRKHSMD--DKVIKDWYELTTPDGIWADMEVSSFCVGSCTGKYFENPAYPPF 496  
Qy 480 -----GSTLD----- 484  
Db 497 TVGSKATSYPVDFVDSNASEWKSIOSSISATAKTSSTSVSSSSSTIDYMTLAPGKGN 556  
Qy 485 -NPYKINNSGVMLPIINKTIPPTAMHYGDIPEYVNHVLEGYLEARVTRALIKL-TEKR 542  
Db 557 NYPPYAIYNMGGSDSLATHAVSPNATHADGTVEYDHNLYGIQENATYHALLEVFNKR 616  
Qy 543 PFVLSRSTFGSGKYTAHTWGDNAATWDLVYSIPSMLDGFLGPIPMVGADICGFLCNTT 602  
Db 617 PFMISRSTFPRAKWTCHWGDNTADWAYAYPSIPQAFSGIAGLPFGADVCGFNGSD 676  
Qy 603 BELCRWIQLGAFYFSPDRHSSIGTYYQELYRWESVAASARKVLGLRYLLPYFYLMYE 662  
Db 677 SELCSRWMLGSGFFPYRNHNYLGAIDQEPYVWESVAEARTSMARIRYLLPPYYTLJHE 736  
Qy 663 AOLNGIPARLPFPSPDDIKTYGISSQFLGKGMVSPVLKPGVSVYATYPRGN---- 718  
Db 737 SHITGLPILRAFSGWQFNDRLSGVONQFFVGGGLVVTPLVPGVDKVKGVFPAGKEEV 796  
Qy 719 WFDLFDYTRSVTASTGRYVTLSPAPDHINVHIOEGNILAMOGKAMTTOARKTPPHLLVV 778  
Db 797 YDWTY-OREVHFQKGNELDAPLGHIPLHIRGGNVLPTEPGYTVAESRQNPFGGLIVA 855  
Qy 779 MSDCGASFGELFLDDGVEVTMGVNRGKMTFVKFIAASAKQTCIITSDVVSGEFAVSQKVV 938  
Db 856 LDNDGKAQGLYLDGGSVLVDSS-----LLVSFVSNDTILSASP--SGDYKADQP-- 904  
Qy 839 IDKVTILGL 847  
Db 905 LANVTILGV 913

## RESULT 15

AAW74090

ID AAW74090 standard; Protein; 1827 AA.

XX AC AAW74090;

	XX	04-MAY-1999	(first entry)	
	DT			
	XX			
	DE			
	XX		Human hSI protein sequence.	
	KW	Gastro-intestinal transport receptor; binding protein; hSI; HPPI;		
	DZH:	hpPPTI; human; GI tract receptor; sucrose-isomaltase complex;		
	KW	intestinal peptide-associated transporter; hypertension; diabetes;		
	KW	osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;		
	KW	therapeutic agent delivery; therapy; probe.		
	OS	Homo sapiens.		
	XX			
	PX	WC9851325-A2.		
	DD			
	PD	19-NOV-1998.		
	XX			
	PF	15-MAY-1998;	98WO-US10088.	
	XX			
	PR	15-MAY-1997;	9TUS-0046595.	
	XX	(CYTO-) CYTOGEN CORP.		
	PA	(ELAN-) ELAN CORP PLC.		
	PA			
	PI	Alvarez VL, Bellinka BA, Cagney GM, Carter JM, Lambkin IJ;		
	PI	Omahony DJ, Patterson CA, Singleton J;		
	XX	MPf; 1999-009368/01.		
	DR			
	XX	New proteins that bind specifically to receptors at the		
	CC	gastro-intestinal tract and related nucleic acid - chimaeras and		
	PT	antibodies, used to deliver therapeutic or diagnostic agents to, or		
	PT	through, the gastrointestinal tract, e.g. insulin or leuprolide		
	XX	Disclosure; Fig 3; 294pp: English.		
	PS	This sequence is the human hSI protein. The invention relates to		
	XX	purified proteins (I) that bind specifically to at least one of the		
	CC	gastro-intestinal (GI) tract receptors human intestinal		
	CC	peptide-associated transporter (HPPI), hpEPRI, DZH and human		
	CC	sucrose-isomaltase complex (hSI). (I) provide active transport of		
	CC	therapeutic agents through human and animal GI tissue (into the blood)		
	CC	for in vivo delivery, particularly for treatment or prevention		
	CC	of hypertension, diabetes, osteoporosis, haemophilia, anaemia, cancer,		
	CC	migraine, or angina pectoris. Specifically they are used to deliver		
	CC	insulin or leuprolide, but many other suitable therapeutic agents are		
	CC	disclosed, including genes or inhibitory nucleic acid, imaging agents and		
	CC	antigens. (II) may also provide targeting to the GI tract. Other uses of		
	CC	(I) are: (i) to determine the level of specified receptors in a sample		
	CC	(in a binding assay); and (ii) to screen for molecules that bind (I).		
	CC	Immunogenic analogues or derivatives of (I) are used to raise antibodies		
	CC	and in immunoassays. The antibodies are used to locate, detect and		
	CC	measure (I), e.g. for imaging, monitoring treatment, tissue analysis		
	CC	etc., also for peptide purification and immobilisation.		
	XX			
	SQ	Sequence	1827 AA;	
		Query Match	27.7%; Score 1318.5; DB 20; Length 1827;	
		Best Local Similarity	35.1%; Pred. No. 1.le-102;	
		Matches	319; Conservative 162; Mismatches 238; Indels 129; Gaps 32	
	QY	35 VIGYGKVKSVKVDS-GTRRSITALPOLAKNSVVGPDIQLLSITASLESDRLRVTG 93		
	DB	109 VNDHGYNVDMMTTTSIGVEAKLRIP-----SPTLFNDINSVLFTQNQTPNFRFKID 164		
	QY	94 AKHRWEIPDNILHRHQPPPPPSSLSLYRTLLSPSTNRKRILLSHNPSDULFSL-IN 152		
	DB	165 PNNRRYEVP-----HQ-----YVKETGTGV-----SDTLVDKVA 195		
	QY	153 TTPFGGITGRSKTHDVLPATDPDPNPNTFLIFTDQYLHTSSLPGTRAHYGLGE--HS 210		
	DB	196 QNFFSIQTVIRKSNGKTLDFTSIGP-----LVYSQYLQISARLPSD--YIVIGISQVHK 247		

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 ; Search time 61.7052 Seconds  
(without alignments)  
3772.184 Million cell updates/sec

Title: US-10-043-418-4

Perfect score: 4751

Sequence: 1 MSLHFWPFIWVVFVFSLS.....EVSKLVLGVKKFEXRLRLT 902

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23:\*

1: sp archaea:\*

2: sp bacteria:\*

3: sp fungi:\*

4: sp human:\*

5: sp invertibrate:\*

6: sp mammal:\*

7: sp mhc:\*

8: sp organelle:\*

9: sp phage:\*

10: sp plant:\*

11: sp rodent:\*

12: sp virus:\*

13: sp vertebrate:\*

14: sp unclassified:\*

15: sp virus:\*

16: sp bacteriap:\*

17: sp archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4751	100.0	902	10 Q9LYF8	Q9lyf8 arabidopsis
2	4746	99.9	902	10 Q22444	Q22444 arabidopsis
3	2578.5	54.3	879	10 Q9JLY2	Q9jly2 hordeum vul
4	2201.5	46.3	910	10 Q8WV9	Q8wv9 pinus pinas
5	2127	44.8	916	10 Q9AVC3	Q9avc3 physcomitre
6	2120.5	44.6	915	10 Q9S7Y7	Q9s7y7 arabidopsis
7	2114.5	44.5	907	10 Q9P2P6	Q9p2p6 arabidopsis
8	2089.5	44.0	928	10 Q9LEC9	Q9lec9 solanum tub
9	2086.5	43.9	929	10 Q9LGC6	Q9lgc6 oryza sativ
10	2049	43.1	935	10 Q9ZP04	Q9zp04 tropaeolum
11	1952	41.1	855	10 Q9LZT7	Q9lzt7 arabidopsis
12	1552	32.7	932	13 Q73626	Q73626 coturnix co
13	1518.5	32.0	937	6 Q9MYM4	Q9mym4 bos taurus
14	1514.5	31.9	952	4 Q8IWE7	Q8iwe7 homo sapien
15	1508	31.7	953	11 Q8SGI6	Q8sgi6 mus musculus
16	1506	31.7	953	11 Q9IZ45	Q9iz45 mus musculus

17	1453.5	30.6	992	3 Q9UV08	Q9uv08 emericella
18	1449	30.5	873	13 Q73632	Q73632 coturnix co
19	1417	29.8	1734	4 Q8TE24	Q8te24 homo sapien
20	1350	28.4	995	3 Q9URX4	Q9urx4 schizosacch
21	1278.5	26.9	955	3 Q9C1S7	Q9c1s7 emericella
22	1272.5	26.8	920	5 Q9NFX8	Q9nfx8 penaeus van
23	1261	26.5	1743	5 Q19004	Q19004 caenorhabdi
24	1187	25.0	955	5 Q21750	Q21750 caenorhabdi
25	1072	22.6	856	5 Q20722	Q20722 caenorhabdi
26	914.5	19.2	728	2 Q9RHZ2	Q9rhzz allicyclobac
27	900	18.9	921	10 Q9FN05	Q9fn05 arabidopsis
28	877.5	18.5	914	4 Q8IZM5	Q8izm5 homo sapien
29	876.5	18.4	925	4 Q8TET4	Q8tet4 homo sapien
30	876	18.4	966	4 Q9P0X0	Q9p0x0 homo sapien
31	875.5	18.4	746	16 Q8XIN9	Q8xin9 clostridium
32	875.5	18.4	914	4 Q8IZM4	Q8izm4 homo sapien
33	868	18.3	763	16 Q9YAE7	Q9yae7 listeria mo
34	864	18.2	763	16 Q9ZP84	Q9zpf8 listeria in
35	862	18.1	944	4 Q14697	Q14697 homo sapien
36	861.5	18.1	779	16 Q8YN00	Q8yn00 anabaena sp
37	855	18.0	751	16 Q8RDL1	Q8rdl1 thermoanaer
38	854.5	18.0	944	11 Q8BHN3	Q8bhn3 mus musculu
39	851.5	17.9	769	4 Q8IWX0	Q8iwx0 homo sapien
40	851.5	17.9	966	11 Q08794	Q08794 mus musculu
41	851	17.9	991	10 Q93Y12	Q93y12 arabidopsis
42	833	17.5	919	10 Q24375	Q24375 solanum tub
43	832	17.5	944	6 P79403	P79403 sus scrofa
44	831.5	17.5	653	6 Q9BE70	Q9be70 macaca fasc
45	820	17.3	818	16 Q8YLG7	Q8y1g7 anabaena sp

#### ALIGNMENTS

#### RESULT 1

Q9LYF8	PRELIMINARY;	PRT;	902 AA.
AC Q9LYF8:			
DT 01-OCT-2000 (Tremblrel. 15, Created)			
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)			
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)			
DE Alpha-glucosidase 1 (AT5G11720/T22P22_110).			
GN T22P22_110.			
OS Arabidopsis thaliana (Mouse-ear cross).			
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.			
OX NCBI_TaxID=3702;			
RN [1]			
RP SEQUENCE FROM N.A.			
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoft A.,			
RA Bancroft I., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.;			
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.			
RN [2]			
RP SEQUENCE FROM N.A.			
RA EU Arabidopsis sequencing project;			
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.			
RN [3]			
RP SEQUENCE FROM N.A.			
RA Shinn P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,			
RA Sanh J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D.,			
RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,			
RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,			
RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,			
RA Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,			
RA Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,			
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;			
RT "Arabidopsis cDNA clones.";			
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.			
RN [4]			
RP SEQUENCE FROM N.A.			
RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,			
RA Chan M.M., Chang C.H., Dale J.M., Hayashizaki Y., Hsuan V.W.,			

```
RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen X., Palm C.J.,
RA Quach H.L., Sakurai T., Satoh M., Seki M., Southwick A., Tang C.C.,
RA Toriumi M., Wallender E.K., Wong C., Wu H.C., Yamada K., Yu G.,
RA Yuan S., Shinzaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis ORF clones";
RJ Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL163814; CAB87690.1; -
DR EMBL: AY053414; AK96644.1; -
DR EMBL: BT002222; AAN72233.1; -
DR InterPro: IPR000322; Glyco_hydro_31.
DR Pfam: PF01055; Glyco_hydro_31; 1.
DR PROSITE: PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE: PS00707; GLYCOSYL_HYDROL_F31_2; 1.
SQ SEQUENCE 902 AA; 101118 MW; D856817DAA00093B CRC64;

Query Match 100.0%; Score 4751; DB 10; Length 902;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 902; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSSLHWPNNIFIVVVFSSRSQVLEEEESTVVGYYVVRSGVDSNRQVLTAKLDLI 60
DB 1 MSSLHWPNNIFIVVVFSSRSQVLEEEESTVVGYYVVRSGVDSNRQVLTAKLDLI 60
QY 61 KPSSVYAPDIKSLNLHVSLETSELRIRITDSSQORWEIPETVIPRAGNHSRRFSTEED 120
DB 61 KPSSVYAPDIKSLNLHVSLETSELRIRITDSSQORWEIPETVIPRAGNHSRRFSTEED 120
QY 121 GGNSENNFLADPSSDLVFTLHNTTPFGFSVSRSSGDIILFDTSPPSSDSNTYFIKQDF 180
DB 121 GGNSENNFLADPSSDLVFTLHNTTPFGFSVSRSSGDIILFDTSPPSSDSNTYFIKQDF 180
QY 181 LQLSSALPENRSLNYGIGETKRSFRLLPGETMTLWNADIGSENPVNLVYGSHPFYMDVR 240
DB 181 LQLSSALPENRSLNYGIGETKRSFRLLPGETMTLWNADIGSENPVNLVYGSHPFYMDVR 240
QY 241 GSKGNEEAGTTHGVLLNSNGMDVKYEGHRTITNVIGVIDLTVFAGSPSEVMNQYTEL 300
DB 241 GSKGNEEAGTTHGVLLNSNGMDVKYEGHRTITNVIGVIDLTVFAGSPSEVMNQYTEL 300
QY 301 IGRPAPMPYMSFGHQCRYGKNSDLEYVVDGYAKAGIPILEVMTDIDYMDGKDFTL 360
DB 301 IGRPAPMPYMSFGHQCRYGKNSDLEYVVDGYAKAGIPILEVMTDIDYMDGKDFTL 360
QY 361 PVNFPEDKMQSFVDTLHKNGQKYYLLDPGIGVDSSYGTYNRGMEADVFIKRNCEPYLGE 420
DB 361 PVNFPEDKMQSFVDTLHKNGQKYYLLDPGIGVDSSYGTYNRGMEADVFIKRNCEPYLGE 420
QY 781 GELFLDGGENLRMGAGGNDWTLVKFCVYVTKSVVLRSEVNVNPEYASQMKWSIGKVTF 840
```

```
DB 781 GELFLDGGENLRMGAGGNDWTLVKFCVYVTKSVVLRSEVNVNPEYASQMKWSIGKVTF 840
QY 841 VGFENVKVTYEVRTSERLRSRISLIKTVDNDPRLFSVEVSKLSLLVGGKFFEMELR 900
DB 841 VGFENVKVTYEVRTSERLRSRISLIKTVDNDPRLFSVEVSKLSLLVGGKFFEMELR 900
QY 901 LT 902
DB 901 LT 902

RESULT 2
O22444 PRELIMINARY; PRT; 902 AA.
AC O22444;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Alpha-glucosidase 1 (EC 3.2.1.20).
GN AGU1
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Landsberg;
RX MEDLINE=98045620; PubMed=9411456;
RA Monroe J.D., Hall B.D., Gough C.M., Stephen A.L.;
RT "Nucleotide sequence of an alpha-glucosidase gene (Accession No.
AF014806) from Arabidopsis thaliana (PCR97-141).";
RL Plant Physiol. 115:863-863 (1997).
DR EMBL: AF014806; AAB82656.1; -
DR InterPro: IPR000322; Glyco_hydro_31.
DR Pfam: PF01055; Glyco_hydro_31; 1.
DR PROSITE: PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE: PS00707; GLYCOSYL_HYDROL_F31_2; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 902 AA; 101106 MW; 71F2AB9F4D8D22EA CRC64;

Query Match 99.9%; Score 4746; DB 10; Length 902;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 901; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSSLHWPNNIFIVVVFSSRSQVLEEEESTVVGYYVVRSGVDSNRQVLTAKLDLI 60
DB 1 MSSLHWPNNIFIVVVFSSRSQVLEEEESTVVGYYVVRSGVDSNRQVLTAKLDLI 60
QY 61 KPSSVYAPDIKSLNLHVSLETSELRIRITDSSQORWEIPETVIPRAGNHSRRFSTEED 120
DB 61 KPSSVYAPDIKSLNLHVSLETSELRIRITDSSQORWEIPETVIPRAGNHSRRFSTEED 120
QY 121 GGNSENNFLADPSSDLVFTLHNTTPFGFSVSRSSGDIILFDTSPPSSDSNTYFIKQDF 180
DB 121 GGNSENNFLADPSSDLVFTLHNTTPFGFSVSRSSGDIILFDTSPPSSDSNTYFIKQDF 180
QY 181 LQLSSALPENRSLNYGIGETKRSFRLLPGETMTLWNADIGSENPVNLVYGSHPFYMDVR 240
DB 181 LQLSSALPENRSLNYGIGETKRSFRLLPGETMTLWNADIGSENPVNLVYGSHPFYMDVR 240
QY 241 GSKGNEEAGTTHGVLLNSNGMDVKYEGHRTITNVIGVIDLTVFAGSPSEVMNQYTEL 300
DB 241 GSKGNEEAGTTHGVLLNSNGMDVKYEGHRTITNVIGVIDLTVFAGSPSEVMNQYTEL 300
QY 301 IGRPAPMPYMSFGHQCRYGKNSDLEYVVDGYAKAGIPILEVMTDIDYMDGKDFTL 360
DB 301 IGRPAPMPYMSFGHQCRYGKNSDLEYVVDGYAKAGIPILEVMTDIDYMDGKDFTL 360
QY 361 PVNFPEDKMQSFVDTLHKNGQKYYLLDPGIGVDSSYGTYNRGMEADVFIKRNCEPYLGE 420
DB 361 PVNFPEDKMQSFVDTLHKNGQKYYLLDPGIGVDSSYGTYNRGMEADVFIKRNCEPYLGE 420
```



```

QY 421 VMPGKYVDFDLNPAATAATWSNEIKMPOBILPLDGLWIDMNELSNFTSPSSGSSLDLP 480
Db 421 VMPGKYVDFDLNPAATAATWSNEIKMPOBILPLDGLWIDMNELSNFTSPSSGSSLDLP 480
QY 481 PYKINNNGDKRPINNKTVPATSIHFGNISEYDAHNLGLLEAKATHQAVVDITGKRPFIL 540
Db 481 PYKINNNGDKRPINNKTVPATSIHFGNISEYDAHNLGLLEAKATHQAVVDITGKRPFIL 540
QY 541 SRSTFVSSGKYTAHWTDGNAAKWEDLAYSI PGILNFGFLGIPMGADICGFSHDTTESLC 600
Db 541 SRSTFVSSGKYTAHWTDGNAAKWEDLAYSI PGILNFGFLGIPMGADICGFSHDTTESLC 600
QY 601 RRMWOLGAYFPFARDHSSLGTAQELYLWDSVASSARKVLGLRMELLPHLYTLMYEAHVS 660
Db 601 RRMWOLGAYFPFARDHSSLGTAQELYLWDSVASSARKVLGLRMELLPHLYTLMYEAHVS 660
QY 661 GNPAPLPFFSPQDTKTYEIDSQFLIGKSI MVSPALKQGA VADAYFPAGNWFDFENYS 720
Db 661 GNPAPLPFFSPQDTKTYEIDSQFLIGKSI MVSPALKQGA VADAYFPAGNWFDFENYS 720
QY 721 FAVGDSGKHVRDLTPADHVNHVREGSVAVMGCEALITRDARKTPYQLLVVASLENIS 780
Db 721 FAVGDSGKHVRDLTPADHVNHVREGSVAVMGCEALITRDARKTPYQLLVVASLENIS 780
QY 781 GELFLDDGLENRMAGGNGNRDWTLVKFCYVTGKSVLSEVNVPEYASOKMWSIGKVTFF 840
Db 781 GELFLDDGLENRMAGGNGNRDWTLVKFCYVTGKSVLSEVNVPEYASOKMWSIGKVTFF 840
QY 841 VGFENVNKTYEVRTSERLSPRISLIKTVDSDNDPRLSVEVSKLSLLVGKKFEMRLR 900
Db 841 VGFENVNKTYEVRTSERLSPRISLIKTVDSDNDPRLSVEVSKLSLLVGKKFEMRLR 900
QY 901 LT 902
Db 901 LT 902

```

## RESULT 3

```

Q9LLY2 PRELIMINARY; PRT; 879 AA.
ID Q9LLY2
AC Q9LLY2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE High pI alpha-glucosidase.
GN AGL97.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Igri;
RX MEDLINE=20267959; PubMed=10806244;
RA Frandsen T.P., Lok F., Mirgorodskaya E., Roepstorff P., Svensson B.;
RT "Purification, enzymatic characterization, and nucleotide sequence of
RT a high-isoelectric-point alpha-glucosidase from barley malt.";
RL Plant Physiol. 123:275-286(2000).
DR EMBL; AF118226; AAF76254.1; -.
DR InterPro; IPR000322; Glyco_hydro_31.
DR InterPro; IPR002052; N6_Mtase.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
DR PROSITE; PS00092; N6_MTASE; 1.
SQ SEQUENCE 879 AA; 96558 MW; AF9235ECE7D15B44 CRC64;

```

```

Query Match 54.3%; Score 2578.5; DB 10; Length 879;
Best Local Similarity 54.6%; Pred. No. 3.4e-176;
Matches 495; Conservative 150; Mismatches 211; Indels 51; Gaps 12;

```

```

QY 12 IVVVFFSLRSSQVLEEEESTVVGYYVRSVGVDSNRQVLTAKLDLIKPSVSYAPDIK 71
Db 7 LLLCLCLCLFAPLCSKKEEGPLAAGYRVSRAVDGGRRLRAEAAATGGASSTGDVQ 66
QY 72 SLNLHVSLETSELRIRITDSSQORWEIPETVIPRAG-----NHSPPRFSTEBDGGNSPE 126
Db 67 RLAVASLETSDSLRVITDADHPREVEPQDIIPRAPADVLHADPASPAPLQG----- 121
QY 127 NNFLADSSDLVTLNHTTPGFSVSRSGDILFDTSPDSSDSNTYFFPKQFOLQSSA 186
Db 122 SRVLSAAGSOLVITVH--ASPFRFTVSRRTGDI LFDTPG-----LVFRDKYLEVTS 173
QY 187 LPENRSNLGIGGHTKRSFRLIPGETMTLWNADIGSENPDVNLGSHPPFVMDVRGSKNE 246
Db 174 LPAGRASLYGLGHTKSSFLRHNDSTFLWNADIGASYVDVNLGSHPPFVMDVR----- 227
QY 247 EAGTTGVLNLLNNGMDVKYEGHRIITYNVIGGVIDLYVFAGSPSPVMNQYTELICRPAP 306
Db 228 PPGTAHGVLLSSNGMDVLYGGSYVTKVIGGVLDYFFAGPNPLAVVDQYTLQIGRPAP 287
QY 307 MPYKSGFCHOCRYGKXVNSDLEVYVDGAKAGIPIEVMMTDIDYMDGYKDPDLPVNFPE 366
Db 288 MPYKSGFCHOCRYGKXVNSDLEVYVDGAKAGIPIEVMMTDIDYMDGYKDPDLPVNFPE 366
QY 367 DKMQSFVDTLHKKGQKYVLIDPGIGVDSYGTNRGMEADVFIRKNGEPYLGGEVWPGKV 426
Db 348 AELRPFVDRLHRNAQKYVLIDPGIRIDATYGTFFVRGMOODIFLKNGTGNFVGNWPGDV 407
QY 427 YFDFLNPAAATFWSNEIKMPOBILPLDGLWIDMNELSNFIT--SPSSGSSLDLPYKIN 485
Db 408 YFDFMHPRAAEFWAREISLFRRTIPVDGLWIDMNEISFNYPEPM--NALDDPYRIN 464
QY 486 NSGDKRPINNKTVPATSIHFGNISEYDAHNLGLLEAKATHQAVVDITGKRPFILSRSTF 545
Db 465 NDTGTRPINNKTVPASAVHYGGVTEYDAHNLGLLEAKATHQAVVDITGKRPFILSRSTF 524
QY 546 VSSGKYTAHWTDGNAAKWEDLAYSI PGILNFGFLGIPMGADICGFSHDTTELCRRWIQ 605
Db 525 VSGRYTAHWTDGNAATGDLRYSI NTMLSFGIFGMPMIGADICGFNGNTTELCGRWIQ 584
QY 606 LGAPYPFARDHSSLGTAQELYLWDSVASSARKVLGLRMELLPHLYTLMYEAHVSQNP 665
Db 585 LGAPYPFSDHSAIFTVRRELYLWPSVAASARKALGLRYQLLPYFTYLMYEAHMTGAP 644
QY 666 RPLFFSPQDTKTYEIDSQFLIGKSI MVSPALKQGA VADAYFPAGNWFDFENYSFVGG 725
Db 645 RPLFFSYPHDVATYGVDRQFLGLGRVLVSPVLEFGATTVDAYFPAGRWYSLYDYLAVAT 704
QY 726 DSGKHVRDLTPADHVNHVREGSVAVMGCEALITRDARKTPYQLLVVASLENISGELFL 785
Db 705 RTGKHVTLPAADTVNVHVAGGTILPLQQSALTTSRARTAFHLLVALAEDGTASGDPLF 764
QY 786 DDGENLRMGAGGNGNRDWTLVKFCYV--TGKSVLSEVNVPEYASOKMWSIGKVTFVGE 844
Db 765 DDGESPEM---GGRSDMVMVRFCEMGSDGAIKVKSEVHNSVAQSRITLVISKVLMDGHR 821
QY 845 NVENYK--TYEVRTSERLSPRISLIKTVDSDNDPRLS-----VEVSKLSLLVGKKF 895
Db 822 SPAAPKLTIVHNSAE-----VEASSAGTRYQNRAGGLGVGHGGLSLVVGEEF 871
QY 896 EMRLRLT 902
Db 872 ELKVAMS 878

```

## RESULT 4

```

Q8VWV9 PRELIMINARY; PRT; 910 AA.
ID Q8VWV9
AC Q8VWV9;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative alpha-xyloidase.

```

GN XY1.1.  
OS Pinus pinaster (Maritime pine).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.  
OX NCBI\_TaxID=71647;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Etolated hypocotyl;  
RA Sanchez M., Gianzo C., Sampedro J., Revilla G., Zarra I.;  
RT "Changes in alpha-xylosidase gene expression during intact and auxin-  
RT induced growth of pine hypocotyls."  
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF448201; AAL40352.1; -;  
DR InterPro; IPR000322; Glyco\_hydro\_31.  
DR Pfam; PF01055; Glyco\_hydro\_31; 1.  
DR PROSITE; PS00129; GLYCOSYL HYDROL\_F31\_1; 1.  
DR PROSITE; PS00707; GLYCOSYL HYDROL\_F31\_2; 1.  
DR PROSITE; PS00038; HLH\_1; 1.  
SQ SEQUENCE 910 AA; 100609 MW; C4B75C7306CC16F9 CRC64;  
  
Query Match 46.3%; Score 2201.5; DB 10; Length 910;  
Best Local Similarity 46.6%; Pred. No. 4.3e-149;  
Matches 439; Conservative 148; Mismatches 247; Indels 109; Gaps 16;  
  
QY 13 VVVFFSLRSQVVLBEESTVVGYYVRSVGVDSNRQ-VLTAKLDLIKPSVYAPDIK 71  
Db 15 VVCAMSLGHGQ-----QVGYG--RLVSDQSGSLGSELIQNTTYGPDIP 63  
  
QY 72 SLNLHVSLETSELRIRITDSSQORWEIPETVPRAGNHSRPFSTEEDGNSPENFLA 131  
Db 64 HRLVYKHETEDRVVRVHTDAQTRKEWVPQELLSR-----EQAPADLPVSSRKA 112  
  
QY 132 DPSSD-----LVTLHTTPFGSVSRSSGDI LFTSDSSDSNTYFTFKQ 179  
Db 113 KPAKSAFEFSKAGGELLIVFISN--PFGFAIKRSGNDVLFSSYGN-----LVFKDQ 164  
  
QY 180 FLOLSALPENRNLVIGIGBHTK-RSPRLIPGETMTLWNAIDGSENPVNLVSHGPFYMD 238  
Db 165 YLEVTTLGLPATAS-LYGLGENTQPGIKILPKAYLTITDTSAINLNTDLYSHGPFYMD 223  
  
QY 239 VRSGKGNEEAGTTHGVLLNSNGMDVKYEGHRITNYVIGVIDLYVPAGSPEMVKNQYT 298  
Db 224 VR-----NGGISRGVLLNSNGMDVFTYNALTYKVIQGVLDYFPFAGTSPLDVVQOYT 277  
  
QY 299 ELIGRPAMPVWSFGHCRGYKYNKVSLEVVVDGYAKAGIPLEVWMTDIDYMGYKDT 358  
Db 278 ALIGRPVAQPTWAFGRFQCRWGYKNVSDITNVVYNYKSQLPLDVWDDHMDGAKDPT 337  
  
QY 359 LDPVNFPECKMQSFVDTLKNGQKYVLI LDPGIGVDSYGTYNRGMADVFIKRNGEPYL 418  
Db 338 LDP-NYPEYKLRPLPDLRIHANGMYVVLIDPGIAINTSYGTQFGWADQVFIKHGSPFL 397  
  
QY 419 GSVWFGKYVPPDFLNPAAATFWSNEIKMFQEI LPLDGLWDMNLSNFTSPUS----- 472  
Db 398 GQVMPGAVYPPDFLNPXTYVFWADEISHPHQMFVVDGLW:DMNEISNFCSGKCSIPTNRS 457  
  
QY 473 -----SGSLDDPPYKINNSGDKRPINKNKTVPATSIHFGNISSEYDAHN 515  
Db 458 CPGTGFWECLDRNTATRWVPPYKINASGTQVPLGFKTIATSSVHNGVLEVDHNS 5:7  
  
QY 516 LYGLLEAKATHQAVVDITKGRPFILSRSTFVSSGKYTAHWGTGDNAAKWEIDLAYSIFGLN 575  
Db 518 LYGLSQAIAATHALQNLDRKPPVLTSTTFVSGSYAAHWTDGNKATWEDLRYISITILN 577  
  
QY 576 FGLFGIPWGNADICGSHDITTEELCRWIOLGAPYFPARDHSSIGTARQELYLWDSVASS 635  
Db 578 FGMFGMPWGNADICGFYDPTTEELCCGRWIOLGAPYFPFSRCHSNLASKRQELYLWDSVAKS 637  
  
QY 636 ARKVLGLRMLRLLPHLYTLMTYEAHVSGMPIARLPFFSPQDKTYEIDSOFLICKSINWSP 695  
Db 638 ARKALGLYRLPLYTLNDAHTTGAPIARLPFFSPQDKTYEIDSOFLICKSINWSP 697

QY 696 ALKQGAIVADYFPAQKWFDLFNYSFAVGDSGKHVRLDTPADHVNTFVREGSIVAMQGE 755  
Db 698 VLYNKTTSVNAVFPKGSWYNLNDMTMAV-KSSGGYVTLQAPMDTINVHVCEGMILPMQRG 756  
  
QY 756 ALITRDARKTPYQLLVV-----ASRLNISGELFLDDGLENLRCMGAGGGRDWTLVKPCRY 810  
Db 757 GMTTIVARMTPTLIIAFPLGFQSTGCKAKGHLFLDSGEDVDMKIAEGKS--TVVDFSAE 814  
  
QY 811 VTGKSVVLRSVNVPEVASKMWSIGKVTFFGFEN-----VENKVTYEV 854  
Db 815 SDGKKVRLVSQVESGSYGLSQGVVVEKMLIILGSKSHLSQIAFQLDGKPTSSSFTYSV 874  
  
QY 855 RFSERLSRPSRLIKTYSNDNDPRFLSVEYSKLSLLVKKFEM 897  
Db 875 QP-----LSTASQSQGGGAIMELNGALPVGRIIDL 906  
  
RESULT 5  
Q9AVC3  
ID Q9AVC3 PRELIMINARY; PRT; 916 AA.  
AC Q9AVC3;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Putative alpha-glucosidase.  
GN PGGLU  
OS Physcomitrella patens subsp. patens.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;  
OC Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.  
OX NCBI\_TaxID=145481;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hiwataishi Y., Nishiyama T., Hasebe M.;  
RT "Establishment of gene- and enhancer-trap systems of the moss,  
RT Physcomitrella patens."  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB057452; BAB39467.1; -;  
DR InterPro; IPR001064; Crystallin.  
DR InterPro; IPR000322; Glyco\_hydro\_31.  
DR Pfam; PF01055; Glyco\_hydro\_31; 1.  
DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; 1.  
DR PROSITE; PS00129; GLYCOSYL HYDROL\_F31\_1; 1.  
DR PROSITE; PS00707; GLYCOSYL HYDROL\_F31\_2; 1.  
SQ SEQUENCE 916 AA; 102282 MW; AA2E2AC6E440D883 CRC64;  
  
Query Match 44.8%; Score 2127; DB 10; Length 916;  
Best Local Similarity 47.4%; Pred. No. 9.7e-144;  
Matches 428; Conservative 150; Mismatches 257; Indels 68; Gaps 17;  
  
QY 36 GYGYVRSVGVDSNRQVLTAKLDLIKPSVYAPDIKSLNLHVSLETSELRIRITDSSQ 95  
Db 31 GAGHRMTSVFEDDGGSGFVANLELITGTEIYGPDISPLRMIAVYDSDRLHVHITDSIHA 90  
  
QY 96 RWEIPETVIPAGNHS-----PRFSTEDGNSPENFLADPPSDLVFTLHTNTTPGFSV 151  
Db 91 RWEVPQDIIPRDPSSSLVTHYKERMHESEGLDPARN---DRQLQLSYTVE---PFGFAI 144  
  
QY 152 SRRSSGDI LFTSP-----DSSD-SNTYFIKQDLQLSSALPENRNLVIGIETHK-RSP 205  
Db 145 TRTSTGECLFNTSPIRODSGEPAPNSVMVFKQQLYEISTQLPRNNS-LFGIGESTRPDGL 203  
  
QY 206 RLIPGETMTLWNAIDGSENPVNLVSHGPFYMDVRSKGNBEEAGTTHGVLLNSNGMDVK 265  
Db 204 RLTRGRLYTLNATDIAAYKVVDVLYGAYFPFYMDIR-----EGCATHGVLNLSNGMDIW 257  
  
QY 266 YEGHRITNVYTGVIDLYVPAGSPFEMVKNQYTELICRPAPMPYWSFGHCRGYKNVS 325  
Db 258 VGEDMLTHVYIGGVLDYFFAGAPLAVIDQYTNLIGRPTMPYWSFGHCRGYETID 317  
  
QY 326 DLEVYVDGYAKAGIPLEVWMTDIDYMGYKDTLDPVNFPECKMQSFVDTLKNGQKYVL 385  
Db 318 EIKDVVYKKNANPLDITMNDIDYMDAYKDTFPDPVRYDENTVREFVKELHANGQOYIV 377

```

QY 386 ILDPGIGVD-SSYGYTYRQGEADVFIKRN-GBPVLGEVMPGKVYFPDFLNPAAATFWGNE 443
Db 378 ILDPGIVGKYNSTLERGLKDDIELKNEFGNYYLAQWPGVYFPDFLHPKSSWNTQE 437
QY 444 IKMFOILPLDGLWDMNLSNFTSPSSGSL-----D 478
Db 438 IADFFKVPFDGLWDMNEASNFTGSGACSFDTLTLMGKNDSDNDRCLLHCVCNGTSRFD 497
QY 479 DPKYKINNSGDKRPINNKVTPATSIHFGNISYDAHNLGLLEAKATHCAQVVDITGKEPF 538
Db 498 DPKYKINHVGTVDNLGVKXTIAMTVRHNGVLEVDYDAHNLGLCESATOKTLDRTVTKGKPF 557
QY 539 ILSRSTFVSSGKYTAHTWTDGNAKMEDLAYSPGILNFGIFGIPMVGADICGFSHDTTTEE 598
Db 558 ILSRSTFVSGAHTAHTWTDGNAKMEDLAYSPGILNFGIFGIPMVGADICGFSHDTTTEE 617
QY 599 LCRNWTQAGAFPPARDHSSLTGAKQELVLDNSVASSARKVJGLRMLLPHLYTLMYBAH 658
Db 618 LCRNWTQAGAFPPARDHSSLTGAKQELVLDNSVASSARKVJGLRMLLPHLYTLMYBAH 677
QY 659 VSGNPIARLPFTSPDPTKTYEIDSDOFLICKIMVSPALKQGAVALDVPAGNMFDFLN 718
Db 678 KSGAPIARLPFTSPDPTKTYEIDSDOFLICKIMVSPALKQGAVALDVPAGNMFDFLN 737
QY 719 YSFAVGDSGKHVRLDTPADHNVHVRGSIIVAMQGEALTTRDARKTPYQLLWV--ASRL 776
Db 738 FSKIV--STGERMLPAPADSNVHVSEGOILPMQEARLTSAEVKTFPTLVVVSADAS 795
QY 777 ENISGELFLDDGLENRMAGGNGRDWTLKPCRYVTGKSVLSEVNVNPEVASOKWSIG 836
Db 796 ASASGKLFVDSGVDI:EMGIQDSS--TFVQFAERSLHSGSLVSRVAGNVALEQGLVLQ 853
QY 837 KYTFVGFEN-VENVKTYEVRTSERLRSRISLIKTVDNDQPRFLSVEVSKLSLVGKKF 895
Db 854 SIRFLGVSGFVSDV-----IVNGER-----IVSAEQLSYDARLESQVLSLJLGRPF 902
QY 896 EMR 898
Db 903 ELR 905

RESULT 6
Q9S7Y7 PRELIMINARY; PRT; 915 AA.
AC Q9S7Y7;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Alpha-xylosidase precursor (ATIG68560/F24J5_10).
GN XYL1 OR F24J5_20.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsie.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RC Sampedro J., Steiro C., Villa T.G., Revilla G., Zarra I.;
RT "Cloning and expression pattern of an alpha-xylosidase gene from
RT Arabidopsis thaliana."
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
RA Lee J.M., Li J., Gonzalez A., Liu K., Vaysberg M., Sakano H.,
RA Chin C., Choi E., Chou J., Alkafi H., Araujo R., Brooks S.,
RA Buehler E., Chao Q., Conn L., Conway A.B., Dunn P., Hansen N.,
RA Howing B., Huizar L., Khan S., Kim C., Palm C., Rowley D., Shinn P.,
RA Walker M., Davis R.W., Ecker J.R., Federpiet N.A., Theologis A.;
RT "The sequence of BAC F24J5 from Arabidopsis thaliana chromosome 1."
RT Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.

```

```

[3]
SEQUENCE FROM N.A.
RP Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Iehida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinzaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
SEQUENCE FROM N.A.
RP Kim C.J., Chen H., Cheuk R., Shinn P., Bowser L., Carninci P.,
RA Chan M.M., Chang C.H., Dale J.M., Hayashizaki Y., Heuan V.W.,
RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA Tang C.C., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G., Yuan S.,
RA Shinzaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis ORF clones."
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF144078; AAD37363.1; -
DR EMBL; AC008075; AAD49987.1; -
DR EMBL; AY057482; AAL0916.1; -
DR EMBL; BT002675; AAO11591.1; -
DR InterPro; IPR000322; Glyco_hydro_31.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
KW SIGNAL.
FT CHAIN 1 27 POTENTIAL.
FT CHAIN 124 915 ALPHA-XYLOSIDASE.
SQ SEQUENCE 915 AA; 102398 MW; 28F9610D8D7EA657 CRC64;

Query Match 44.8%; Score 2120.5; DB 10; Length 915;
Best Local Similarity 44.8%; Pred. No. 2.8e-143;
Matches 422; Conservative 166; Mismatches 277; Indels 77; Gaps 15;

QY 2 SSLHWPENIIVVVVFPFLSRSSQVLEEEESTVGVYVYVSVGVDSNRQVLTAKLDLIK 61
Db 6 SSLAFSLSLALILCFSPQSY-----KTIGKGYRLVSI-BESPDGFGYGLQVQK 56
QY 62 PSSVYAPDIKSLMLHVSLETSELRIRITDSSQORWEIPETVIPR-----AGNHSP 112
Db 57 KNKIYSDITTLRLFKVGHETDSRLRVHITDAQORWEVPYNLLPREQPPQGVKICKSRK 116
QY 113 RRFSTEEDGNSPENNPLADPSSDLVFTLHNTTPGFSVSRSSGDLIFDTPSSDSNT 172
Db 117 SPITVQEISG-----SELIFS-YTTPDFTFAVKRRSNHETLENTT-----S 156
QY 173 YFIPQDFLOLSSALPENRSLNLYGIGETHK-RSFRLIPGETMTLWNADIGSENPVNLG 231
Db 157 SLVFPKQYLSISLSLPKAS-LYGLGNSQANGIKLVNPEPTLYTETDSAINMTDLYG 215
QY 232 SHPFTYMDVRSGKNEEAGTTHGVLLNSNGMDVKYEGHRITNYNVIQGVLDLYVAGPSPE 291
Db 216 SHPMYMDLRNVGCKAYA--HAVLLNSNGMDVYFRGDSLTLYKVGVEDFPYIAGPSPL 272
QY 292 MYMYQYTELIGRPAPMPYWSFGHQRYGKYNVSDLEYVVDGYAKAGIPLEVWMTDIDYM 351
Db 273 NVVDQYTLQIGRPAPMPYWSLGLFHCRCRWGYNHLSVVEDVVDNKKAKIPLDVTWDDHM 332
QY 352 DGYKDFTLDPNFPEDKMQSFVDTLHKNGOKYVLLDPLGIVDSSVGYTVNRGMEADVFIK 411
Db 333 DGHKDFLTNPVAYPRAKLFLKIKHGMKYIIVNDPGIGVNNASGTFORAADAADVFIK 392
QY 412 RNGEPYLGVEWPGKYFPDFLNPAAATFMSNEMFQELPLDGLWDMNLSNFTS-- 469
Db 393 YEGKFLAQWPGVYFPDFLNPFTKTSWMDGEIKRFDLVPIDGLWDMNEVSNFCSGLC 452
QY 470 -----PLSSG-----SSLDDPPYKINNSGDKRPINNKVTPATSIHFGN 507

```

[illegible]



```
Db 421 LNPNGVSWWIDEVRFDLVPDGLWIDMNEASFCTGKCEIPTTHLCLPLNTTTPWCC 480
Qy 473 -----SGSLDDPPPKYKNSGDKRPINKTKVTPATSIHFGNISEYDAHNLGLLEAKATH 526
Db 481 LDCNLTNRWDDEPPPKYKINASGQTARLGENTATSIATHYNGILEYNAHSLYFSQAIATH 540
Qy 527 QAVVDITGRPFILSRSTFVSSGKYTAHWTGDNAAKWEIDLAYSIPGILNFGFLGIPMWGA 586
Db 541 QALQGIQGRPFILSRSTFVSGYAAHWTGDKNGTWNELRYSISTMLNFGIFGPMWGA 600
Qy 587 DICGFSHDTTEELCRRWIOLGAFYFPARDHSSLGTAQELYLWDSVASSARKVLGLRML 646
Db 601 DICGYPQTEELCNRWIELGAFYFPRDHFANFASPRQELYVWESVAKSARNALGNRYL 660
Qy 647 LPHLYTLMEAHVGNPIARPLFFSPQDTKYEIDSOFLIGKSIWVSFALQGAUVA 706
Db 661 LPVLYTLNQAHLTGAPVAPVFFSPDFTPCYGLSTQVLLGASVMSVLEQATSVSA 720
Qy 707 YFPAGNWFDLFNSFAVGDSGKHVRLDTPADHVNHVHREGSIVAMQGEALTTRDARKTP 766
Db 721 MFPPGWNLFDTTKVVUSGEGAVKLDAPLNEINHVFTQNTILPMQRGTTISKEARATP 780
Qy 767 YQLLVV-----ASLENISGELFDDGNGLRMGAGGNRDWTLVKFCYVTGKSVLVRSEV 822
Db 781 FTLVAFPPFCATEAE-AEGAVYVDDDERPEMWLAEOGA--TYVRFVATVRKAVTVRSEV 837
Qy 823 VNPYASAKMWSIGKTVFCFE-----NVENVKTYEVRISERLRSRISLIKTVPD- 873
Db 838 ELGSLQKGLLEKLSVLGLETGRDLAVHVDGANATATSRPYFAGAEALHGHDRV 897
Qy 874 NDDPRFLSVEVSKLSLLVGLKKFEM 897
Db 898 EGHKKSVMVEVGLALPLGKSFTM 921

RESULT 10
Q9ZP04 PRELIMINARY; PRT; 935 AA.
AC Q9ZP04;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Alpha-D-xylosidase precursor.
CS Tropaeolum majus (Common nasturtium).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Tropaeolaceae; Tropaeolum.
OX NCBI_TaxID=4020;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv, tall climbing mixed; TISSUE=Cotyledon;
RA Crombie H.J., Chengappa S., Jarman C., Sidebottom C., Reid G.;
RT "Molecular characterisation of a xyloglucan oligosaccharide-acting
RT alpha-D-xylosidase from the cotyledons of nasturtium (Tropaeolum majus
RT L) seeds.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ131520; CRAI0382.2; -
DR InterPro: IPR000322; Glyco_hydro_31.
DR Pfam: PF01055; Glyco_hydro_31; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
KW SIGNAL.
FT SIGNAL.
FT CHAIN 1 24 POTENTIAL.
FT CHAIN 25 935 ALPHA-D-XYLOSIDASE.
SQ SEQUENCE 935 AA; 104937 MW; 22DE6901E9CE19BD CRC64;

Query Match 43.1%; Score 2049; DB 10; Length 935;
Best Local Similarity 45.1%; Pred. No. 4e-138;
Matches 411; Conservative 150; Mismatches 291; Indels 60; Gaps 13;

Qy 33 TVVGYGVVRSVGVDSNRQVLTAKLDLRKSSVYAPDIKSLNHLVSLTSELRIRITOS 92
Db 39 TKIGKGRYLSIETPDGGFL-GHLOVKQPNKIYGADIPLLQLYVKHESQDRLRVHITDA 97
```

```
Qy 93 SQQWEIPETVI PRAGNHSRRFSTEDGGNSPENFLA----DPSDLVFTLHNTTTPGF 149
Db 98 EKQWEVYNNLPR--EQPP---VVEANDRBIPOKNLITVSEISGSELIFSYRPRDRFGF 152
Qy 150 SVSRSSGDIILFDTSPSSSNTYFIKQDFLOQLSSALPENRSLNYGIGETHK-RSRPLI 208
Db 153 AVKRSNGETLFNSSSDPSDFGEMVFKQYLEISTKLPKDAS-LYGLGENTQPHGKILY 211
Qy 209 ZGETMTLWADIGSENPDVNLGSHPPYMDVRGSKNEEAGTTHGVLLLSNGMDVKYEG 268
Db 212 PNDPYTLTMDVSAINLADSYGSHPMYDLRNVGGEAYA---HAVLLLSNGMDVFEYRG 268
Qy 269 HRTVTVIGVIDLYVYFAGSPVMVMNQYTELICRPAPMPYMSFGHOCYGYKNVSDLE 328
Db 269 DSLTYKIIGGVDFYFFIGFAPLDVDOYTAFIGRPAPMPYMSLGFHQCRWGYHNLVIE 329
Qy 329 YVVDGYAKAGILEPVMVTDDIDYMDGYKDFTLDPVNFPEDKMQSFVDTLHKNGQKYVIL 388
Db 329 DVERYKNAKIPLDVINWDDHMDGKDFLTNTKQYRPPQLLAFELDKSIGIMKYIVID 388
Qy 389 PGIGVDSYGYTNRGMEADVFIRKNGEPYLGEMVPGKYFPDFLNPAAATPWSNEIKMFQ 448
Db 389 PGIAVNSYGYTQRLANDVFIRYEGEPFLAQWPGAVNPFDFLNPXTVDWMGDEVRRFH 448
Qy 449 EILPLDGLWIDMNSLNF-----TSPLSSG-----SSLDDPPYKI 484
Db 449 ELVPVGLWIDMNEDELEFLFWMENPOGKQCTGEGGWCCLCKNITKTRWDEPPYKI 508
Qy 485 NNSGDKRPINKTKVTPATSIHFGNISEYDAHNLGLLEAKATHQAVVDITGRKPFILSRST 544
Db 509 NATGVAQPIGFKTIATSCTHYNGVLEYDAHSIYGFSTVATHKALQLEGKRPFILTRST 568
Qy 545 FVSSGKYTAHWTGDNAAKWEIDLAYSIPGILNFGFLGIPMWGADICGFSHDTTEELCRWI 604
Db 569 FVSGHYAAHWTGDNQGTWENLRYSISTMLNFGIFGVPWVGSDICGYPQTEELCNRWI 628
Qy 605 QLGAFYFPARDHSSLGTAQELYLWDSVASSARKVLGRMLRLLPHLYTLMEAHVSGNPI 664
Db 629 EVGAFYFPRDHFANFASPRQELYOWESVASSARNALGNRYKLLPFLYTLTEAHIRAPI 688
Qy 665 ARPLFFSPQDTKYEIDSOFLIGKSIWVSFALQGAUVAVDAYFPAGNWFDFLNFYSFVG 724
Db 689 ARPLFFTPNYPECYGVSTQFLGSSLMISPVLEQKTEVKALFPFGTWSLDMTETVD 748
Qy 725 GDSKXHVRLDTPADHVNHVHREGSIVAMQGEALTTRDARKTPYQLLV---VASLENISG 781
Db 749 SKGQYVTLDAPLHVNVHLYQNTILPMQOGLLSKEARMTPTFLITVFPAGATDGOAKG 808
Qy 782 ELFLDDGLENLRMGAGGNRDWTLVKFCYVTGKSVLVRSEVNPENYASAKMWSIGKTVFV 841
Db 809 NLFLDKDELPEMKLNGYS--TYVEFYATLNQGAVKVWSQVQEGKFDALDNCWSIEKVTVL 866
Qy 842 GFEN-----VENVKTYEVRISERLRSRISLIKTVSNDDPRFLSVEVSKLSL 889
Db 867 GLSNKQVGSLEIDGSPVSGISKVEMSSSE-----QIFVGKLDABESKPSLMVVEVKGLDI 921
Qy 890 LVGKKFENLRPL 901
Db 922 PVGNFVNSWKN 933

RESULT 11
Q9LZT7 PRELIMINARY; PRT; 855 AA.
AC Q9LZT7;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 95.9 kDa protein.
GN F16L2.150.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
```



```
Db 278 ARDV-AFTESNLYGAHPFYLLM-----EEGDAHGCVFLNLSNAMEVALQAPGLTMT 330
QY 276 IGGVLDLYVAGSPSPMMQYTELGRPAPMPYWGFGHCRCRYGYNVSDLEYVVDGVA 335
Db 331 IGGVLDLYVAGSPSPMMQYTELGRPAPMPYWGFGHCRCRYGYNVSDLEYVVDGVA 335
QY 336 KAGIPLEVMWTDIDYMDGYKDFLDPVNFPEDKMQQFVDTLHKNQKQYVLIIDPGIGVDS 395
Db 391 NFQIPQDAQNNDIDYMDGYKDFLDPVNFPEDKMQQFVDTLHKNQKQYVLIIDPGIGVDS 448
QY 396 ---SYGTYNRGMADVIK-RNGEPYLGVMQGVFPDFLPNPAATFWSNEIKMFQEL 451
Db 449 PRGSYMPFDEGLRRGLFLNTTQGTLLIGQVPGYTAYPDPSNTDTHQWLENLQRFTHV 508
QY 452 PLGLWTDNMLNFIITSPSSGS-----SLDDPYK---INNSGDKRPIPNKTVPAT 501
Db 509 PFDGLWTDNMLNFIITSPSSGS-----DSEEGCPGELDSPYTPAVLGNS-----LTAKTVCAS 558
QY 502 -----SIHFGNISEYDAHNLVGLLEAKATHQAVVD:TKRPPFILSRSTFVSSGKYTAHWT 556
Db 559 AEQNASHVH-----YNLHNLVGLKEAATASALIRIGKPPFVISRSTFVSSGKYTAHWT 612
QY 557 GDNAAKWEIDLAYSIGILNFGILGIPMWGADICGFSHDTTELCRRWIOLGAPYPPARDH 616
Db 613 GDNRSQWKMYISIPGMLSPFLGIPLVGADICGFSGSTSEELCTRWMLGAFYPPSRNH 672
QY 617 SSLGTAQELVLDNSVASSARK-VLGLRMLLPHLYTLMYEAHVSNPPIARPLFESFPD 675
Db 673 NNQNEKAQDTATSPSARTAKMDALLTRYSLLPFLTLFRAHQOETVARPUFFFPMD 732
QY 676 TKTYEIDSQFLIGKSIWVSPALQKQAVADAYFPAGNWFOLFNYFSAVGDSGSKHVRD 735
Db 733 VATYGLDQRLWGQSLVTPVLEPGADSVLGYFPQGVYDYFTGSSV--NSSGEMLKLSA 790
QY 736 PADHNVHVRECSIWAMQGEALTRDARKTPYOLLVVASRLENIISGELFDDGENL 791
Db 791 PLDHLNLHREGSILPTQKPGITSKATRGNPLHLIALSTRATAWGDMDDGESL 846

RESULT 13
Q9MYM4 PRELIMINARY; PRT; 937 AA.
AC Q9MYM4
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Acidic alpha-glucosidase.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
CX NCBI TaxID=9913;
RN 1
RP SEQUENCE FROM N.A.
RX MEDLINE=20188772; PubMed=10723725;
RA Dennis J.A., Moran C., Healy P.J.;
RT "The bovine alpha-glucosidase gene: coding region, genomic structure,
RT and mutations that cause bovine generalized glycogenosis.";
RL Mamm. Genome 11:206-212(2000).
DR EMBL; AF171666; AAF81636.1; -
DR EMBL; AF171665; AAF81636.1; -
DR HSP; P01359; 2ESP.
DR InterPro; IPR000519; P_trefol.
DR Pfam; PF01055; Glyco_hydro_31.
DR Pfam; PF00088; trefol; 1.
DR SMART; SM00018; PD; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
DR PROSITE; PS00025; P_trefol; 1.
DE SEQUENCE 937 AA; 104756 MW; B7E9A2D66BAF5026 CRC64;
```

```
Query Match 32.0%; Score 1518.5; DB 6; Length 937;
Best Local Similarity 38.3%; Pred. No. 5.1e-100;
Matches 336; Conservative 136; Mismatches 230; Indels 115; Gaps 24;

QY 54 TAKLDLIKPSVYAPDIKSLNLHVSLETSELRIRITDSSQORWEIP-ETVIPRAGHSP 112
Db 136 TATUTRAVP-TFFPKDINTLRDLMLMETESRLHFTIKDPANRRYEVLET--PRVYSQAP 192
QY 113 RRFSTEEDGNGSPENFLADPSSDLVFLHNT-----TPPGFSVSRSSGDLIDTSDSS 168
Db 193 -----FTLYSEFSEEPFGVVRKLDGRVLLNTTV--- 223
QY 169 DSNTYFIKQDQLSSALPENRNLXIGIHTKRSFRLIPGETMTLWNADIENSEDNV 228
Db 224 ---APLFAQDFLOLSTSLPS--QHITGLAEHLGSLMLSTNWKITLWNRDIAPE-PNVN 277
QY 229 LYGHPPFMDVGRSGKNEEAGTHGVLLNLSNGMDVKYE-GHRITYNVIGVLDLYVFAG 287
Db 278 LYGHPPFVLV-----EDGGLAHGVFLNLSNAMDVVLQPSPALSMRSTGGLIDVYIFLG 331
QY 288 PSEPMWNOYTELIGRPAPMPYWSFGFHQCRGYGYNVSDLEYVVDGYAKAGIPLVWMTD 347
Db 332 PEKSVVQOYLDVVGYPFMPYPYWGFLHLCRWGYSTSAITRQVVVENMTRAYFFLDVOMND 391
QY 348 IDYMDGYKDFLDPVNFPEDKMQQFVDTLHKNQKQYVLIIDPGI---GVDSSYGTYNRGM 404
Db 392 LDYMDARRDFTFNKDHFGD--FPAMVQELHQGRRYIMIVDPAISSSGPAGTYRYPDEGL 449
QY 405 EADYFI-KRNGEPYLGVMQGVFPDFLPNPAATFWSNEIKMFQELPLDGLWIDNML 463
Db 450 RRGVFIITNETQPIIGQVMPGLTAPDPFTNPTLDMQDMVTEFHAQVFDGMMIDNRP 509
QY 464 SNFTTSPSS--GSSLDPPYKINNSGDKRPIPNKTVPATSIHFGNISEYDAHNLVGLLE 521
Db 510 SNFVSGVDGCPDNLNPPYLPQVVG--TLRAATICASHQFELS-THYDLNLVGLTE 566
QY 522 AKATHQAVVDITGRPFILSRSTFVSSGKYTAHWTGDNAAKWEIDLAYSIGILNFGILFI 581
Db 567 ALASHRALVKARGMRPFVISRSTFAGHGRYSGHWTGDMWNSWEQLSVPEILLFNLLGV 626
QY 582 PMVGADICGESHDTTEELCRWIOLGAFYFPARDHSSIGTARQELVLDNSVASSA-RKVL 640
Db 627 PLVGADICGFLGNTSEELCRWIOLGAFYFPARNHNLNSQOEYRFSFETAQQAQMKAF 686
QY 641 GLRMELLPHYLTYLMEARVSNPIARPLFFSFPODTKYEIDSQFLIGKSIWVSPALQKQ 700
Db 687 TLRVLLPYLYTLFRAHVRGETVARPLFEFFPEDPSTWTVDRQLLWGEALLITPVLEAE 746
QY 701 AVAVDAYFPAGNWFOLFNYFSAVG-----DSGKHVRDTPADHNVHVHVRG 747
Db 747 KVEVTGYFPQGTWYDLOQTPVMEAFGSLPPAPLTSVIHSKQGWVTLSPALDTINVLHLAG 806
QY 748 SIVAMQGEALTRDARKTPYOLLVVASRLENIISGELFDDGENLRMGAGGNRDWTLVKF 807
Db 807 HIIPMQGPALTTTSRKHQMALAVALTASGEAQGLFWDGSGLVLDGG---DYTQLIF 863
QY 808 RCYVTGKSVLLRSEVNPVPEYASKMKWSIGKTVFVGFE-----NVNENKTYEVRSERLRS 863
Db 864 -----LAKNNTFVNKLHVSVSEGASLQLRNVTVLGVAT-----AP 898
QY 864 RISLIKTV-----SDNDDPRFLSVEVSKLSLVGKKF 895
Db 899 QOVLCNSVPVSNFTFSPDTETLAIPV---SLTWGEQF 932

RESULT 14
Q8IWE7 PRELIMINARY; PRT; 952 AA.
ID Q8IWE7
AC Q8IWE7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Glucosidase, alpha, acid (pompe disease, glycogen storage disease type
```





```

QY 631 SVASSA-RKVLGLRMRLPHLYTMVEAHVSGNPIARPLFFSPQDTKYEIDSOFLICK 689
Db 689 ETACQAMRKAFALRYALLPYLYTLFRAHVRGDTVARPLFLEPPDPSTWSVDRQLLMGP 748
QY 690 SIMVSPALKOGAVAVDAYFPAGNWFLENYSFAVGG-----DSGKHVRL 733
Db 749 ALLITPVLEPGKTEVTGYFPKGTWYNQMVSVDLSGLTLPSPSSASSFRSAVQSKGWLTL 808
QY 734 DTPADHVNHVREGSIVAMQGEALTTTRDARKTPYQLLVVASRLNIGSELFLLDDGLENLM 793
Db 809 EAPLDTINVHLREGYIIPLOQPSLTTTESRKQPMALAVALTASGEADGELFWDDGESLAV 968
QY 794 GAGGNRDWTLVKFCYVTGCKSVLRSEVNVPEYASKMWSIGKVTFGFENV-ENVKTY 852
Db 869 LERGA---YTLVTFS--AKNNTIYNKLVRTKEGA---ELQUREVTVLGVATAPTOVLSN 920
QY 853 EVRTSERLRSRISLIKTVDSDNDPRFLSVEVSKLSLLVGKKFEM 897
Db 921 GIPVSNFTYSP-----DNKSLAIPV---SILMGELFQI 950

```

Search completed: October 27, 2003, 10:31:06  
Job time : 65.7052 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 ; Search time 12.5417 Seconds  
(without alignments)  
3382.161 Million cell updates/sec

Title: US-10-043-418-4

Perfect score: 4751

Sequence: 1 MSSLHWFNFIFVWVFFSL.....EVSKLSLLVGKKFEMRLRLT 902

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2830.5	59.6	903	1	AGLU SPIOL
2	2719.5	57.2	913	1	AGLU BETVU
3	2451	51.6	877	1	AGLU HORVU
4	1520.5	32.0	952	1	LYAG HUMAN
5	1500.5	31.6	864	1	AGLU MUCJA
6	1471	31.0	953	1	LYAG MOUSE
7	1423	30.0	985	1	AGLU ASPNG
8	1421	29.9	1856	1	MGA_HUMAN
9	1413	29.7	969	1	AGLU SCHPO
10	1393.5	29.3	946	1	AMYG CANAL
11	1360	28.6	1826	1	SUIS HUMAN
12	1358.5	28.6	1826	1	SUIS RABIT
13	1357	28.6	985	1	AGLU ASPOR
14	1347	28.4	958	1	AMYG DEBOC
15	1345.5	28.3	993	1	YAJI SCHPO
16	1322.5	27.8	1812	1	SUIS SUMMU
17	1310	27.6	1840	1	SUIS RAT
18	1271.5	26.8	1070	1	AGLU CANTS
19	1046.5	22.0	923	1	AGLU TETPY
20	1008.5	21.2	787	1	AGL2 BACTQ
21	819	17.2	954	1	YB79 YEAST
22	713.5	15.0	693	1	AGLU SULSO
23	514.5	10.8	731	1	XYLS SULSO
24	512.5	10.8	772	1	YICI ECOLI
25	309	6.5	678	1	YIHO ECOLI
26	225.5	4.7	529	1	YCR2 ERWHE
27	189	4.0	696	1	SP15 TORCA
28	149.5	3.1	972	1	HGBB HAEDU
29	147	3.1	1047	1	CARB THEAC
30	144	3.0	972	1	HGBA HAEDU
31	139	2.9	2334	1	WAPA BACSU
32	126	2.7	1645	1	OMPB RCTYU
33	124	2.6	861	1	GLGB SOLTU

34 123.5 2.6 3343 1 YOG7 CAEEL  
35 123 2.6 1151 1 ITAI\_HUMAN  
36 120 2.5 682 1 E13E\_BACCI  
37 120 2.5 1592 1 GTF2\_STRDO  
38 119.5 2.5 1900 1 STT4\_YEAST  
39 119 2.5 457 1 MURC\_CLOPE  
40 117.5 2.5 733 1 AGAL\_PBDPE  
41 116.5 2.5 1357 1 YJ03\_YEAST  
42 114.5 2.4 1849 1 IGA4\_HAEIN  
43 114 2.4 1694 1 IGA2\_HAEIN  
44 114 2.4 1702 1 IGA2\_HAEIN  
45 113 2.4 555 1 FGMU\_ASPOR

#### ALIGNMENTS

RESULT 1  
AGLU SPIOL  
ID\_AGLU SPIOL STANDARD; PRT; 903 AA.  
AC 004893:  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase)..  
OS Spinacia oleracea (Spinach).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.  
OX NCBI\_TaxID=3562;  
[1]  
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=cv. Dash;  
RX MEDLINE=97238484; PubMed=9132069;  
RA Sugimoto M., Furui S., Suzuki Y.;  
RT "Molecular cloning and characterization of a cDNA encoding alpha-glucosidase from spinach";  
RL Plant Mol. Biol. 33:765-768(1997).  
CC -!- FUNCTION: ALPHA-GLUCOSIDASE I AND II HAVE HIGH ACTIVITY TOWARDS MALTO-OLIGOSACCHARIDES AND STARCH, WHILE FORM III AND IV HAVE HIGH ACTIVITY TOWARDS MALTO-OLIGOSACCHARIDES BUT LOW ACTIVITY TOWARD STARCH.  
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-linked D-glucose residues with release of D-glucose.  
CC -!- PTM: FOUR DIFFERENT FORMS (I-IV) MAY BE PRODUCED BY POST-TRANSLATIONAL MODIFICATION.  
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
-----  
EMBL; D86624; BAA19924.1; --  
PIR; T09143; T09143.  
InterPro; IPR000322; Glyco hydro 31.  
Pfam; PF01055; Glyco hydro 31.  
PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.  
Hydrolase; Glycosidase; Glycoprotein; Signal.  
SIGNAL 1 24  
CHAIN 25 903  
FT ACT\_SITE 465 465  
FT CARBOHYD 64 64  
FT CARBOHYD 219 219  
FT CARBOHYD 400 400  
FT CARBOHYD 491 491  
FT CARBOHYD 501 501  
FT CARBOHYD 600 600  
FT CARBOHYD 903 903  
FT CARBOHYD 100880 MW; 5B054E27C20EC33A CRC64;  
SQ SEQUENCE 903 AA; 100880 MW; 5B054E27C20EC33A CRC64;

Query Match		59.68;	Score 2830.5;	DB 1;	Length 903;
Best Local Similarity		59.67;	Pred. No. 1.8e-183;		
Matches 540;		Conservative 138;	Mismatches 205;	Indels 23;	Gaps 8;
Qy	8	PNIFVIVVFFSLR---	SSQVLEEBESTVVGYGIVVRSVGVDS-NRQVLTAKLDLIKPS	63	
Db	6	PSLALGILLVFLQYL	VAGISTSENDEPGYIGYKVKSVKVDGSTRSITALPOLVKN	65	
Qy	64	SVYAPDIKSNLHVLS	ETSERLRIRITDSSQQRWEIPETVPRAGNHSP-----RRFST	117	
Db	66	SVYGPDIQLLSITAS	LESNDRLRVITDAKHRSWEIPDNILHR---HQPPPPPHSLSL	122	
Qy	118	EEGSGSPENN----	FLADPSSDLVFTLHNTTPGFSVSRSSGJILFDTPSSDSNTY	173	
Db	123	YRTLSSPTNRRKILL	SHPNDSUFTSCINTTPGFTISRKSTHVDLFDATPPTNPFT	192	
Qy	174	FIFKQFLOLSSALP	ENRSLNYGIGETKRSFRILPGETMTLWNAOTGSENPDVNLVYGS	233	
Db	183	LIFIDQVHLTSSLP	GTTRAHYIGLGEHSKPTFQLAHNQTLTMRADIPSSNPVNLVYGS	242	
Qy	234	PFYMDVSGKNEBAG	THGVLLNSNGMDVKYEGHRIYNNVIGGVLDLVFVAGPSPMW	293	
Db	243	PFYMDVRS---	PVAGSTHGVLLNSNGMDVEYTGNRITYKVGIIIDLYFFAGSPGV	299	
Qy	294	MNOYTELIGRPAMP	YPSFGPHOCRYGKVDLEYVVDGYAKAGIPLEVYMTDIDYDG	353	
Db	300	VEQTRVIGRPAMPY	WAFGQQCRYGHDVYELQSVVAGYAKAGIPLEVYMTDIDYDA	359	
Qy	354	YKDTLDPVNFPMQ	SFVDTLHKNQKYVLLDIPGIVDSSYGTYNRMEADVFIKN	413	
Db	360	YKDTLDPVNFPMQ	KKFNVLHKNQKYVLLDIPGISTNKTYETVIRGMKHDPVLRN	419	
Qy	414	GEPLGVWPKYVFP	DLPAAPATFWSNETKMPQELPLDGLWIDMVELSNITPSLS	473	
Db	420	GKPYLGSVWPKYV	FPDPLKPSALTFTWDEIKFENLLPVDGLWIDMVELSNITPS	479	
Qy	474	GSSLDPPYKINNS	GDKRPINNKVTPATSIHFGNISYDAHNLVGLLEAKATHQAVVDIT	533	
Db	480	GSTLDNPPYKINNS	GVMLPIINKTIPTAMHYGDIPBYNVNLPFYLEARVTRAAIKLT	539	
Qy	534	GKRPFLISRSTFV	SSGKYTAHTWTDGNAAKMEDLAYSIPGILNFGLPIMPVGADICOF	593	
Db	540	EKRPFVLSRSTF	SSGKYTAHTWTDGNAATNDLVYSIPSMFLDGLFIMPVGADICGFL	599	
Qy	594	DTTELCRRWTLQ	CAFPFAPDHSLSGTARQELVLDVSVASSARKVLGLMRLLPHLYTL	653	
Db	600	NTTELCRRWTLQ	CAFPFAPDHSLSGTARQELVLDVSVASSARKVLGLMRLLPHLYTL	659	
Qy	654	MYEARVSGNPIAR	PLFFSFPDDTKTYEIDSOFLIGKIMVSPALKQCAVAVDVFAGNW	713	
Db	660	MYEAQNLNGIP	PIARPLFFSFPDDTKTYGISSOFLIGKGMVSPVLKPGVSVTAFFRGNW	719	
Qy	714	PCLNYFAVGGSG	KGVRLDTPADHNVNVRBGSIVAMOGREALTTDRARKTPVQCLLWA	773	
Db	720	PDLFDYTRSVTAS	TGRVYVTLSPADPHNVHFIQENILAMQOKAMTTOAARKTPFHLVVM	779	
Qy	774	SRLENISELFLD	GENLRWMCAGGGRDWTLVKFCYVYTGKSVVLRSEVNVPEVASKKW	833	
Db	780	SDCASFGEULFD	GDGVEVTMGVNRGK--WTFVKPIAASAKOTCIITSDVVGSGFAVSQW	837	
Qy	834	SIGKVTVGFEN	VNKTYEYTRSERLSRPSRISLIKTYSNDNDRFLSVESVSKLSLLVGK	893	
Db	838	VIDKVTILGLK	KGKINGYTVTCAVTRKDGKSKLKTTPDRKG-EFIVAEISGLNLLGR	896	
Qy	894	KFEMRL	899		
Db	897	EFKLVL	902		

RESULT 2

AGLU\_BETVU

ID\_AGLU\_BETVU STANDARD; PRT; 913 AA.

AC	004911;				
DT	15-JUL-1998	(Rel. 36, Created)			
DT	15-JUL-1998	(Rel. 36, Last sequence update)			
DT	30-MAY-2000	(Rel. 39, Last annotation update)			
DE	Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase).				
OS	Beta vulgaris	(Sugar beet).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;				
OC	Caryophyllidae; Caryophyllales; Chenopodiaceae; Beta.				
OX	NCBI_TaxID=161934;				
RN	[1]				
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				
RC	STRAIN=CV. NK-152;				
RX	MEDLINE=97321863; PubMed=9178565;				
RA	Matsui H., Iwanami S., Ito H., Mori H., Honma M., Chiba S.;				
RT	"Cloning and sequencing of a cDNA encoding alpha-glucosidase from sugar beet.";				
RL	Biosci. Biotechnol. Biochem. 61:875-880(1997).				
RN	[2]				
RP	ACTIVE SITE, AND SEQUENCE OF 464-472.				
RX	MEDLINE=95252592; PubMed=7766184;				
RA	Iwanami S., Matsui H., Kimura A., Ito H., Mori H., Honma M., Chiba S.;				
RT	"Chemical modification and amino acid sequence of active site in sugar beet alpha-glucosidase.";				
RL	Biosci. Biotechnol. Biochem. 59:459-463(1995).				
CC	!- FUNCTION: HIGH ACTIVITY FOR ALPHA-GLUCAN.				
CC	!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-linked D-glucose residues with release of D-glucose.				
CC	!- PTM: THE N-TERMINUS IS BLOCKED.				
CC	!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
DR	EMBL; D89615; BAA20343.1; -.				
DR	PIR; JC5463; JC5463.				
DR	InterPro; IPR00322; Glyco_hydro_31.				
DR	Pfam; PF01055; Glyco_hydro_31.1_				
DR	PROSITE; PS00129; GLYCOSYL_HYDROL_F31.1; 1.				
DR	PROSITE; PS00707; GLYCOSYL_HYDROL_F31.2; 1.				
KW	Hydrolase; Glycosidase; Glycoprotein; Signal.				
FT	SIGNAL	1	28	POTENTIAL.	
FT	CHAIN	29	913	ALPHA-GLUCOSIDASE.	
FT	ACT SITE	469	469		
FT	CARBOHYD	54	54	N-LINKED (GLCNAC. .) (POTENTIAL).	
FT	CARBOHYD	404	404	N-LINKED (GLCNAC. .) (POTENTIAL).	
FT	CARBOHYD	495	495	N-LINKED (GLCNAC. .) (POTENTIAL).	
FT	CARBOHYD	517	517	N-LINKED (GLCNAC. .) (POTENTIAL).	
FT	CARBOHYD	728	728	N-LINKED (GLCNAC. .) (POTENTIAL).	
FT	CARBOHYD	823	823	N-LINKED (GLCNAC. .) (POTENTIAL).	
SQ	SEQUENCE	913	AA;	102117 MW; 02ACAF0B505369CC CRC64;	
Query Match					
Best Local Similarity 57.2%; Score 2719.5; DB 1; Length 913;					
Matches 516; Conservative 146; Mismatches 219; Indels 25; Gaps 7;					
Qy	8	PNIFIV-VVVFSSRSQV	LEESTVVGIVVRSVGVDS-SNRQVLTAKLDLIKPS	64	
Db	12	PTLAVVLVLVCMV	VEGATTSKNQNGEAGVGYQVKNKVDNSTGKSLTALLQIRNSP	71	
Qy	65	VYAPDIKSNLHVLS	ETSERLRIRITDSSQQRWEIPETVPR-----AGNHSP	113	
Db	72	VYGPDIHFLSFTAS	FEEDDTLRIFTDANNRWEIPNEVLPRPPPPPLSSIQHLPK	131	
Qy	114	RFSTEEDGNSPENN	FLADPSSDLVFTLHNTTPGFSVSRSSGJILFDTPSSDSNTY	173	
Db	132	PIPQ-----NQPT	TVLSHPHSDLAFTLPHFTTFFGFTYRKSTHVDLFDATPISNPTTF	186	

```
QY 174 FIFKQFQLOLSSALPENRNLNGIGETKSRPLIPGETMTLWNADIGSENPDVNLGSH 233
D 187 LIYKQYQLOLSSALPAQQAHLGLGHEHTKPTQLAHNQILTLWNADIASFNRDLNLYGSH 246
QY 234 PYMDVRSGKNEEAGTHGVLLNSNGMDVKYEGHRTYNNVIGVLDLYVPAGSPENV 293
D 247 PYMDVRSS--PMVGSTHGVLLNSNGMDVYTGDRITYKVGIGIIDLXYFAGRTPEMV 303
QY 294 MNQYTELGRPAMPVWSPGFHQCVRGYKNVSDLEVVDGKAGIPLEVMTDIDYMDG 353
D 304 LDQYTKLGRPAMPVWSPGFHQCVRGYKNVSDLEVVDGKAGIPLEVMTDIDYMDG 363
QY 354 YKDFLDPVNFPEDMQSVFDTLHKNQKQYVLDLPDGVDSYSGYNNRMEADVFYKRN 413
D 364 PKDFLDPVNFPEDMQSVFDTLHKNQKQYVLDLPDGVDSYSGYNNRMEADVFYKRN 423
QY 414 GPYIGEVWPKGVYFDFLNPAAATFWSNEIKMFQELPLDGLWIDMELNSFNITSPUSS 473
D 424 GNPYLGSMVPGVYVDFLNPAAATFWSNEIKMFQELPLDGLWIDMELNSFNITSPUSS 483
QY 474 GSSLDPPVKINNSGDKPPINNKTYPATSIHFGNISEYDAHNLGLLEAKATHQAVVDIT 533
D 484 GSTLNDPPVKINNSGDKPPINNKTYPATSIHFGNISEYDAHNLGLLEAKATHQAVVDIT 543
QY 534 KGRPFILSRSTFVSSGKYTAHTGDNAAKWEGLAYSIPIGLNFGFGIPMVGADICGFESH 593
D 544 TRGPFILSRSTFVSSGKYTAHTGDNAAKWEGLAYSIPIGLNFGFGIPMVGADICGFESH 603
QY 594 DTTELCLRWIOLGAFYFPPARHSSLGTAQBYLWDSVASSARKVLGRKRLPLHLYTL 653
D 604 STTEELCLRWIOLGAFYFPPARHSSLGTAQBYLWDSVASSARKVLGRKRLPLHLYTL 663
QY 654 MYEAHVGNGPIARPLFFGFPODTKYEIDSOFLGCKSMVSPALKOGAVAVDAYFPAGNW 713
D 664 MYDANLRGSPARPLFFGFPODTKYEIDSOFLGCKSMVSPALKOGAVAVDAYFPAGNW 723
QY 714 FDLFNYSFVAGDSDGKHVRLDTFADHNVNVRGSIIVAMQGEALTRDARKTPYOLLVVA 773
D 724 VLSNYSVSSVSAGTVSVLSAPPDINHVIHEGNIVAMQGEALTRDARKTPYOLLVVA 783
QY 774 SRLENISGELFDGLENLMGAGGNRWDTLVKPCYVTKSVLRSEVNPPEYASKMKW 833
D 784 SDHVAAGTGLFDGLENLMGAGGNRWDTLVKPCYVTKSVLRSEVNPPEYASKMKW 841
QY 834 SIGKTYFVGFENNVKTYEVRTSERLSRPSRLIKTVSDNDPRLSVESYKSLLVGK 893
D 842 VMDKITILGKRVIKETVQKDAGAIKVGGLGRRTSHNOGGFFVSV-ISDLRLQVLGQ 900
QY 894 KPEML 899
D 901 AFKLEL 906
RESULT 3
AGLU HORVU STANDARD; PRT; 877 AA.
AC Q43763;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase).
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poaceae;
OC Tricaceae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RC STRAIN=cv. Morex; TISSUE=Aleurone;
RX MEDLINE=96178863; PubMed=8616248;
RA Tibbot B.K., Skadsen R.W.;
RT "Molecular cloning and characterization of a gibberellin-inducible,
```

```
RT putative alpha-glucosidase gene from barley.";
RL Plant Mol. Biol. 30:229-241 (1996).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-
CC linked D-glucose residues with release of D-glucose.
CC -!- TISSUE SPECIFICITY: HIGH LEVELS SEEN IN THE ALEURONE AND SCUTELLUM
CC AFTER GERMINATION, WHILE LOW LEVELS ARE FOUND IN DEVELOPING SEEDS.
CC -!- DEVELOPMENTAL STAGE: LEVELS INCREASE STEADILY THROUGHOUT
CC IMBIBITION REACHING MAXIMUM LEVELS AT DAY 7. DURING GERMINATION,
CC LEVELS INCREASE FROM DAY 2, REACH MAXIMUM LEVELS AT DAY 3 AND
CC DECLINE AFTER DAY 5.
CC -!- INDUCTION: BY GIBBERELLIN A3 (GA).
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC ENBL; U22450; AAB02985.1; --
CC PIR; S65057; S65057.
CC InterPro; IPR000322; Glyco_hydro_31.
CC Pfam; PF01055; Glyco_hydro_31; 1.
CC PROSITE; PS00129; GLYCOSYL HYDROL_F31_1; 1.
CC PROSITE; PS00707; GLYCOSYL HYDROL_F31_2; 1.
CC Hydrolase; Glycosidase; Glycoprotein; Signal.
CC SIGNAL 1 23
CC FT CHAIN 24 877 ALPHA-GLUCOSIDASE.
CC ACT SITE 437 437 BY SIMILARITY.
CC FT CARBOHYD 191 191 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 298 298 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 338 338 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 391 391 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 471 471 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 570 570 N-LINKED (GLCNAC. .) (POTENTIAL).
CC SQ SEQUENCE 877 AA; 96933 MW; E3B5C16E4588C492 CRC64;
Query Match 51.68; Score 2451; DB 1; Length 877;
Best Local Similarity 55.28; Pred. No. 7.9e-158;
Matches 471; Conservative 136; Mismatches 191; Indels 56; Gaps 13;
QY 69 DIKSLNLHVSLETSELRIRITDSSQORWEIPETVIPRAG-----NHSRRFSTEEDGNN 123
D 59 DVQELAVYASLETDSRLRVITDADHPREVPQDILPRPAPGVDLHDAPPASSAPLOG-- 116
QY 124 SPENFIADPSDLVFTLHNTTFFGFSVSRSSGDIILFDTSPPSSDSNTYFIKDOFLQL 183
D 117 ---RVLSPAGSDLVLTVH--ASPFRTVSRSSGDTLFDTPAG-----LVERDKYLEV 164
QY 184 SSALPENRNLNGIGETKSRPLIPGETMTLWNADIGSENPDVNLGSHPFYMDVRGSK 243
D 165 TSALPAGRSLYGLGHEHTKSRPLIPGETMTLWNADIGSENPDVNLGSHPFYMDVRA-- 222
QY 244 GNEEAGTHGVLLNSNGMDVKYEGHRTYNNVIGVLDLYVPAGSPENVMTQYTELIGR 303
D 223 ---PGTAHGVLLNSNGMDVKYEGSVYTKYIGVLDLYVPAGSPENVMTQYTELIGR 278
QY 304 PAPMPYMSFGHQCVRGYKNVSDLEVVDGKAGIPLEVMTDIDYMDGKFTDLPVN 363
D 279 PAPMPYMSFGHQCVRGYKNVSDLEVVDGKAGIPLEVMTDIDYMDGKFTDLPVN 338
QY 364 FPEDMQSGVDTLHKNQKQYVLDLPDGVDSYSGYNNRMEADVFYKRNPEYLG 420
D 339 FTAELRPFVDRLHRNAQKQYVLDLPDGVDSYSGYNNRMEADVFYKRNPEYLG 398
QY 421 VMPGKYFDPFLNPAATFWSNEIKMFQELPLDGLWIDMELNSFNIT-SPLSSGSLDD 479
D 399 VMPGKYFDPFLNPAATFWSNEIKMFQELPLDGLWIDMELNSFNIT-SPLSSGSLDD 455
QY 480 PPYKINNSGDKRPINNKTYPATSIHFGNISEYDAHNLGLLEAKATHQAVVDITGRPFI 539
D 480 PPYKINNSGDKRPINNKTYPATSIHFGNISEYDAHNLGLLEAKATHQAVVDITGRPFI 539
```

456 PPXRNNDGTRPNNKTVRPLAVHYGVVTEYBEHNLFLGLEARATGRGYLDRGRKFFV 515  
 540 LSRSTFVSSGKYTAHWTDGNAAKWEDLAYSGILNFGLEGI PMWGADICGFSHDTTEEL 599  
 516 LSRSTFVSGRYTAHWTDGNAATWGLDLYSINTLMSGLFGMPMGADICGFGNGTTEEL 575  
 600 CRRRIQLGAFYPPFARDHSSLGCTARQELYLWDSVASSARKVGLGRMLLPHLYTLMYEAHV 659  
 576 CGRWIQLGAFYPPFSDHSAFTVRELYLWPSVAASORKALGRYQLLPFYFTLMYEAHM 635  
 660 SGNFIAPLFFSPQDQTKTVEIDSOFLIGKIMVSPALKGAVADAYFPAGWFDLFNY 719  
 636 TGAPIARPLFFSPDHVATYQDFQLGRGLVSPVLEPGPTTVDAYFPAGRWYRLDY 695  
 720 SFAYGDSGXHVRLDTADHNVHVHVGSSIVAMGGEALTTTRDARKTPYQLLVVASLENI 779  
 696 SLAVATRTGKHVRUPAPADTVNHTGTGLPQQSALTTSRRKTAFLHLLVALAEDGTA 755  
 780 SGELFLDDGLENLRMGAGGGRNDTLVKFRVCYVTGK--SVVLSEVVNPEYASRMKMSIGK 837  
 756 SGYLFLLDDGDSPEYGR--RSDMSVRFNKYIPNNKGAIKVKSEVVHNSYAQSRTLVISK 812  
 838 VTFVGFVENVK--TVEVTSELRSPRLIKTVSDNDPRFLS-----VEVSKUS 888  
 813 VVLGHRSPAPKXLTVHNSAE-----VEASSAGTRYQNAGGLGVGHIGGLS 862  
 889 LLVGKKEFMRLRT 902  
 863 LVGGEFELKVMAS 876

RESULT 4  
 LYAG HUMAN STANDARD: PRT: 952 AA.  
 AC P10253; Q14351; Q16302;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Lysosomal alpha-glucosidase precursor (EC 3.2.1.20) (Acid maltase).  
 GN GAA.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP 230-249; 332-345; 349-370; 394-409; 480-513; 520-545; 703-719; 726-731;  
 RP AND 795-803.  
 RC TISSUE=Placenta, Testis, and Urine;  
 RX MEDLINE=89005058; PubMed=3049072;  
 RA Hoefsloot L.H., Hoogeveen-Westerveld M., Kroos M.A., van Beunum J.,  
 Reuser A.J.J., Oostra B.A.;  
 RT "Primary structure and processing of lysosomal alpha-glucosidase;  
 RT homology with the intestinal sucrase-isomaltase complex.";  
 RL EMO J. 7:1697-1704(1988).  
 RN [2]  
 RP REVISIONS.  
 RA Reuser A.J.J.;  
 RL Submitted (JUN-1990) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90262651; PubMed=2111708;  
 RA Martinik F., Mehler M., Tzall S., Meredith G., Hirschhorn R.;  
 RT "Sequence of the cDNA and 5'-flanking region for human acid alpha-  
 RT glucosidase, detection of an intron in the 5' untranslated leader  
 RT sequence, definition of 18-bp polymorphisms, and differences with  
 RT previous cDNA and amino acid sequences.";  
 RL DNA Cell Biol. 9:85-94(1990).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91097465; PubMed=2268276;  
 RA Hoefsloot L.H., Hoogeveen-Westerveld M., Reuser A.J.J., Oostra B.A.;  
 RT "Characterization of the human lysosomal alpha-glucosidase gene.";

Biochem. J. 272:493-497(1990).  
 [5]  
 RN ACTIVE SITE.  
 RX MEDLINE=91310614; PubMed=1856189;  
 RA Hermans M.M.P., Kroos M.A., van Beunum J., Oostra B.A.,  
 Reuser A.J.J.;  
 RT "Human lysosomal alpha-glucosidase. Characterization of the catalytic  
 RT site.";  
 RL J. Biol. Chem. 266:13507-13512(1991).  
 RN [6]  
 RP CARBOHYDRATE-LINKAGE SITES.  
 RX MEDLINE=93168114; PubMed=8435067;  
 RA Hermans M.M.P., Wisselaar H.A., Kroos M.A., Oostra B.A.,  
 Reuser A.J.J.;  
 RT "Human lysosomal alpha-glucosidase: functional characterization of  
 RT the glycosylation sites.";  
 RL Biochem. J. 289:681-686(1993).  
 RN [7]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=95327152; PubMed=7603530;  
 RA Reuser A.J.J., Kroos M.A., Hermans M.M.P., Bijvoet A.G.A.,  
 Verbeet M.P., van Diggelen O.P., Kleijer W.J., van der Ploeg A.T.;  
 RT "Glycogenosis type II (acid maltase deficiency).";  
 RL Muscle Nerve 3:S61-S69(1995).  
 RN [8]  
 RP VARIANT ASN-91.  
 RX MEDLINE=90365036; PubMed=2203258;  
 RA Martinik F., Bodkin M., Tzall S., Hirschhorn R.;  
 RT "Identification of the base-pair substitution responsible for a human  
 RT acid alpha glucosidase allele with lower 'affinity' for glycogen (GAA  
 RT 2) and transient gene expression in deficient cells.";  
 RL Am. J. Hum. Genet. 47:440-445(1990).  
 RN [9]  
 RP VARIANT GSD-II THR-318.  
 RX MEDLINE=91353580; PubMed=1652892;  
 RA Zhong N., Martinik F., Tzall S., Hirschhorn R.;  
 RT "Identification of a missense mutation in one allele of a patient  
 RT with Pompe disease, and use of endonuclease digestion of  
 RT PCR-amplified RNA to demonstrate lack of mRNA expression from the  
 RT second allele.";  
 RL Am. J. Hum. Genet. 49:635-645(1991).  
 RN [10]  
 RP VARIANT GSD-II LYS-521.  
 RX MEDLINE=91379015; PubMed=1898413;  
 RA Hermans M.M.P., de Graaff E., Kroos M.A., Wisselaar H.A., Oostra B.A.,  
 Reuser A.J.J.;  
 RT "Identification of a point mutation in the human lysosomal alpha-  
 RT glucosidase gene causing infantile glycogenosis type II.";  
 RL Biochem. Biophys. Res. Commun. 179:919-926(1991).  
 RN [11]  
 RP VARIANTS GSD-II ARG-643 AND TRP-725.  
 RX MEDLINE=94004908; PubMed=8401535;  
 RA Hermans M.M.P., Kroos M.A., de Graaff E., Oostra B.A., Reuser A.J.J.;  
 RT "Two mutations affecting the transport and maturation of lysosomal  
 RT alpha-glucosidase in an adult case of glycogen storage disease type  
 RT II.";  
 RL Hum. Mutat. 2:268-273(1993).  
 RN [12]  
 RP VARIANT GSD-II GLU-645, AND VARIANTS ILE-816 AND ILE-927.  
 RX MEDLINE=93168115; PubMed=8094613;  
 RA Hermans M.M.P., de Graaff E., Kroos M.A., Wisselaar H.A.,  
 Willemsen R., Oostra B.A., Reuser A.J.J.;  
 RT "The conservative substitution Asp-645-->Glu in lysosomal alpha-  
 RT glucosidase affects transport and phosphorylation of the enzyme in an  
 RT adult patient with glycogen-storage disease type II.";  
 RL Biochem. J. 289:687-693(1993).  
 RN [13]  
 RP VARIANT GSD-II GLU-645, AND VARIANTS ILE-816 AND ILE-927.  
 RX MEDLINE=92096118; PubMed=1684505;  
 RA Martinik F., Mehler M., Bodkin M., Tzall S., Hirschhorn R.,  
 Hirschhorn R.;  
 RT "Identification of a missense mutation in an adult-onset patient with  
 RT glycogenosis type II expressing only one allele.";



164 TKVPD-ANIYGEVTV-APFRTHNVT--TLWARD-----NPPDFYRNIIYGAHPYQEVVRD 216  
QY 242 SKGNEEAGTHGVLLNSNGMDVKYEGHRITTVNIGGVIDLTVFAGPS--PEMVMMQOYTE 299  
Db 217 -----GKAHGALLMNAHGMDVITTEGRIYKVIIGIILDFVFPAPKSGKPNDSLSTAYTD 269  
QY 300 LIGRAPMPYWSFGFHQCRGYKNYNSDSLBYVDGVGAKAPILEVMMTDDIDYMDGYKDFTL 359  
Db 270 LIGKPMMPGHMLGMHRCRYGPNIDKVTAKRYKEANIPLQTVWVDIDYMEETKDFTF 329  
QY 360 DPVNFPEDKQKQFVDTLHKNGQKLYLIDPGICVDSSYGYTVNRGEADVFIKR-NGEPEYL 418  
Db 330 DKVNFQDPMI GLGEQLHKDGGVYVMDVPAISANTTYPYVRGTEMDVWIKNAQSDGFI 389  
QY 419 GEVWPGKVYFPDFLNPAAATFWSNEIKMFQOEILPLDGLWIDMNLNSFITSPUSSG--- 474  
Db 390 GSVWPGFTTFPDWHPNATKYNNKEIIDFVMDLGVDSLWIDMNEPASFCGLSCGSGKVDA 449  
QY 475 -----SSLDPPP-----YKINNSGDKRPINNKTVP 499  
Db 450 GNQPYRWYTYEBEQAAHNRWEKELKAMGNPGEERNLLYPKAIINNGAGN--LSEFTVA 507  
QY 500 AFSIHFCNITSEVDANHLVGLLEAKATHQAVVDITGK-RPFIILSRSTFVSGKYTAHWITGD 558  
Db 508 TTAHLVGNIPHYDIHNLVGHAEHSHTQALIKHKIRKIRPVILTRSRFPGSGKSVGHWITGD 567  
QY 559 NAAKWEDLAYSIPGIINFLGFIPIWVGADICGFSHDITTEELCRRWITQLGAFYFPFARDHSS 618  
Db 568 KHSFPWYLNKSNITANILNFMFGVSYSGADVCGPNSDTTEELCTRWEIGAFYFPFARNHN 627  
QY 619 LGTARQLYLMDSVASSARKVLGLRMRLPLHLYTLMYEARVSGNPIARLPFFSPDQTKT 678  
Db 628 NAAKQOEPYLWESATAESRIATNRYEMLPYFYTLFEESNRLGLGVWRPLIFEPYAYEEL 687  
QY 679 YEIDSQFLGCKSTMVSPALKOGAVADAYEPAGNWFEDLFNYSFAVGDSGKHVR---LDT 735  
Db 688 VSNVDQTLVGSDDLSPVLDEGKTSYKAQPPGQWYDMYTHETVDNKSNKKVKTVLDA 747  
QY 736 PADHVNHVHREGSIVAMQGEALITRDARTKPYQLLYVVASRLENISGELFLDDGENLRMGA 795  
Db 748 PLTHIPHIHGGALIFTKPKYTVGETFATPYNLVIALDKKGQASGRUYIDDDGESLEVK 807  
QY 796 GCGN-----RDWT 803  
Db 808 SSGYTHFHLQWS 819  
  
RESULT 6  
LYAG MOUSE STANDARD; PRT; 953 AA.  
ID LYAG MOUSE AC P70639;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Lysoosomal alpha-glucosidase precursor (EC 3.2.1.20) (Acid maltase).  
GN GAA.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/c; TISSUE=Liver;  
RA Ding J.H., Yang B.Z., Reuser A.J.J., Roe C.R.;  
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
RC -!- FUNCTION: ESSENTIAL FOR THE DEGRADATION OF GLYCOGEN TO GLUCOSE IN  
CC LYOSOMES.  
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-  
CC linked D-glucose residues with release of D-glucose.  
CC -!- SUBCELLULAR LOCATION: Lysosomal.  
CC -!- SIMILARITY: Contains 1 P-type (trefoil) domain.  
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.  
CC







RL EMBO J. 6:2891-2896(1987).  
 CC -!- FUNCTION: MAY SERVE AS AN ALTERNATE PATHWAY FOR STARCH DIGESTION  
 CC WHEN LUMINAL ALPHA-AMYLASE ACTIVITY IS REDUCED BECAUSE OF  
 CC IMMATUREITY OR MALNUTRITION. MAY PLAY A UNIQUE ROLE IN THE  
 CC DIGESTION OF MALTED DIETARY OLIGOSACCHARIDES USED IN FOOD  
 CC MANUFACTURING.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-  
 CC linked D-glucose residues with release of D-glucose.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-  
 CC glucose residues successively from non-reducing ends of the chains  
 CC with release of beta-D-glucose.  
 CC -!- SUBUNIT: Monomer.  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Brush border.  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN SMALL INTESTINE, GRANULOCYTE, AND  
 CC KIDNEY BUT NOT IN SALIVARY GLAND OR PANCREAS.  
 CC -!- PTM: N- AND O-GLYCOSYLATED.  
 CC -!- PTM: DOES NOT UNDERGO INTRACELLULAR OR EXTRACELLULAR PROTEOLYTIC  
 CC CLEAVAGE.  
 CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.  
 CC -!- SIMILARITY: Contains 2 P-type (trefoil) domains.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 CC EMBL: AF016833; AAC39568.2; -.  
 CC Genbank: HGNC:7043; MGAM.  
 CC MIM: 154360; -.  
 CC GO: GO:0005983; P:starch catabolism; TAS.  
 CC InterPro: IPR000322; Glyco\_hydro\_31.  
 CC InterPro: IPR000519; P\_trefoil.  
 CC Pfam: PF01055; Glyco\_hydro\_31; 2.  
 CC Pfam: PF00088; trefoil; 2.  
 CC SMART: SM00018; PD; 2.  
 CC PROSITE: PS00129; GLYCOSYL\_HYDROL\_F31\_1; 2.  
 CC PROSITE: PS00025; P\_TREFOIL; 1.  
 CC Multicatalon enzyme; Transmembrane; Glycoprotein; Hydrolase;  
 CC Glycosidase; Repeat; Signal-anchor; Sulfation.  
 CC INIT MET 0  
 CC DOMAIN 1 12 CYTOPLASMIC (POTENTIAL).  
 CC TRANSMEM 13 33 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 CC (POTENTIAL).  
 CC DOMAIN 34 1856 LUMENAL (POTENTIAL).  
 CC DOMAIN 37 83 SER/THR-RICH.  
 CC DOMAIN 88 132 P-TYPE 1.  
 CC DOMAIN 197 914 MALTASE.  
 CC DOMAIN 952 998 P-TYPE 2.  
 CC DOMAIN 1066 1812 GLUCOAMYLASE.  
 CC ACT\_SITE 528 528 BY SIMILARITY.  
 CC ACT\_SITE 1419 1419 BY SIMILARITY.  
 CC DISULFID 89 117 BY SIMILARITY.  
 CC DISULFID 100 116 BY SIMILARITY.  
 CC DISULFID 111 129 BY SIMILARITY.  
 CC DISULFID 965 982 BY SIMILARITY.  
 CC DISULFID 977 995 BY SIMILARITY.  
 CC MOD\_RES 415 415 SULFATION (POTENTIAL).  
 CC MOD\_RES 424 424 SULFATION (POTENTIAL).  
 CC MOD\_RES 1281 1281 SULFATION (POTENTIAL).  
 CC CARBOHYD 134 134 N-LINKED (GLCNAC. . .)  
 CC CARBOHYD 294 294 N-LINKED (GLCNAC. . .)  
 CC CARBOHYD 456 456 N-LINKED (GLCNAC. . .)  
 CC CARBOHYD 457 457 N-LINKED (GLCNAC. . .)  
 CC CARBOHYD 478 478 N-LINKED (GLCNAC. . .)  
 CC CARBOHYD 706 706 N-LINKED (GLCNAC. . .)  
 CC CARBOHYD 748 748 N-LINKED (GLCNAC. . .)  
 CC CARBOHYD 826 826 N-LINKED (GLCNAC. . .)  
 CC CARBOHYD 884 884 N-LINKED (GLCNAC. . .)  
 CC CARBOHYD 911 911 N-LINKED (GLCNAC. . .)

FT CARBOHYD 976 976 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 988 988 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1254 1254 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1322 1322 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1363 1363 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1387 1387 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1602 1602 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1671 1671 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1841 1841 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1846 1846 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1856 AA; 209721 MW; 81E7AA0CABAB007D CRC64;  
 Query Match 29.9%; Score 1421; DB 1; Length 1856;  
 Best Local Similarity 36.2%; Pred. No. 1e-87;  
 Matches 323; Conservative 157; Mismatches 302; Indels 110; Gaps 27;  
 QY 37 YGYVVRGVGVDNRQVLTAKLDIKSSVYAPDIKSLNLHVSLETSELRIRITDSSQQR 96  
 DB HSHVHEGNLVNTNAG-PTARLKNLPSPVFGSNVDNVLATYQTSNRPHFKLTDQNNR 193  
 QY 97 WEIPETVIPRAGNHSPPRFSTBEDGGNSPENNFADSPDVLFTLH-NTTPFGFVSRRS 155  
 DB 194 FEVPEHV-----QSFSGNAAASLTQVEISROPPFPIKVTTRS 231  
 QY 156 SGDIPTDPTSDSDSNITYFIKQDFQLQSALPENRSLNLYGIGEHKRSFRL-IPGETMT 214  
 DB 232 NNRVLPDSSIGP-----LLFADQFLQLSTRFSS--TNVYGLGEHVHQYRHDNMKMTW 283  
 QY 215 LWNADIGSEPNVNLGSHFHYMDVRGSKNEAGITHGVLLNSNGMDVYKE-GHRTY 273  
 DB 284 IFNRDTPNGNGNLXGAQTFELCL-----EDASGUSFGVFLMNSNAMEVVLQAPAP 338  
 QY 274 NVIGGVIDLVFAGSPPEVMYMNQYTELIGRAPMPYMSFGHOCRYGYKNVSDLEYVDG 333  
 DB 339 RTIGGILDVYFLGNTPEQVQVELELIGRALPSYWALGFHLRSRYEYGTLDNREVER 398  
 QY 334 YAKAGIFLEVMTDIDYMGYKQFTLDPVN--FPEDKMSQFVDTLHKNGQKVVLIIDPG 390  
 DB 399 NRAAQLPVDVQHADIDYMDERRDFTYDSVDFKGFPE-----FVNELNHNGOKLVIIDPA 453  
 QY 391 IGVDS-----YGTNRGMEADVFIKRG--EPYLGVEWPGKVFPPDPLNPAATFWSNEI 444  
 DB 454 ISNNSSSKPYGPDYDRGDMKIMWNSDSDGYTPLIGVEWPGQTVFPDPTNPNCAVWTKF 513  
 QY 445 KMFEIPLDGLWIDMNELSNFITSPLS---SGSLDDPPPYKINNSGDKRPINN----KT 498  
 DB 514 ELFNQVEFDGIWIDMNEVSNFVDSVSGSGCTNNLNPPF-----TPRILDGVLCKTL 567  
 QY 499 PATSI-HFGNISEYDAHNLXGLLEAKATHQAVDI-TGKRPFILSRSTFVSSGKYTAHT 556  
 DB 568 CMDAVQHWG--KQYDIHNLGYSMATAEAAKTVPFNKRSFILTRSTFAGSGKFAAHL 625  
 QY 557 GDNAKWEIDLAYSIPGILNFCGLFIPMVGADICGFSHDTTEELCRRLIOLGAFYPEAR 616  
 DB 626 GDNTATWDDLRSIPGVLEFNLFGIPMVGPDICGFDLTPEELCRRLWMLGAFYFERNH 685  
 QY 617 SSLGTARQE--LYLWDS-VASSARKVLGLRMLRLPHLYTLMYEAHVSGNPIARLPFSFP 673  
 DB 686 NGQYKQDPPASFGADSLNLSRHYLNIRYTLPLLYTLFFRAHSGDVTVARLLHEFY 745  
 QY 674 QDTKYIDISQFLIGKIMVSPALKQGAVADAYFPAGNMFDLPNYSFVAGDGGKHVR- 732  
 DB 746 EDNSTMDVHQGLWPGGLITPVLDEGAEKVMAVYDPAVMYDY-----ETGSQVRW 796  
 QY 733 -----LDTPADHNVHVRREGSIVAMQGEALTTDARKTPYQLLVASRLNISELFLD 786  
 DB 797 RKQVEMELPGDKIGLHRLGYYIFPTQPNNTTLASRKNPLGLIADENKEAKGELFWD 856  
 QY 787 DGENLWAGAGGGRDWTLVKFRVYTGKSVV--LRSEVNVPEYASKMKSIGKVTYVGF 843  
 DB 857 DGETKDTVA---NKVYLLCEP-----SVTONRLEVINISQSTYKDPNNLAFNEIKILGT 906  
 QY 844 ENVENVKTYEVRTSERLSRSPRISLIKTVSDNDPRFLSVEVSKLSLLVGKXF 895



```
Db 331 NITEIMVQRNVIDADIPVETMSDIDYMEKYRDFTVDPVSVSKSDMQTFESOLVSNQH 390
Qy 383 YVLILDPGI-----GVDSSYCTNRMGEADVFKR-NGEPLVGEVWPGKYVFPDFLNP 434
Db 391 YPFIIDAAIYAANPNYHTDSDYPPYAGVEXKIDFLKNPNNGSIYIGAVWPGFTAFFDFNP 450
Qy 435 AAATFWS-----NEIKMF--QEILPLDGLMIDMNELSNF-----466
Db 451 DVVDYWKDCILNITYAFSGNGTVFPFSGIWTDMNEPSSFCVSGSAMIDLNPAEPLTGIS 510
Qy 467 -----ITSP-----LSSGSSLDSDPPY 482
Db 511 KOYSIEPEGVNSVNTYESSAYSASLSNYATATSSVFQIVSPTATPLGLKPDYINIDWPY 570
Qy 483 KINNSGDKRPINNKTV-PATSIHFNGISYDAHNLVGLLEAKATHOAVVDIT-GKRPFIL 540
Db 571 AINNEQNHDIANHIVSPNATHDGT-QRYDIFNMVGYGETKVSVAALTOISNERPFIL 629
Qy 541 SRSTFVSSGKYTAHTWTDGNAKWEDLAYSIPIGLNFGIFPMVGADICGFSHDTEELC 600
Db 630 SRSTFLGSGVYGAHGLGDNHLSNMFFSISGMIVFNMMGIPMVGADVCGFLGDSDEELC 689
Qy 601 RMIQLGAFPPARDHSSGTARQELYLWDSVASSARKVLGLMRLPHLYTLMEAHVS 660
Db 690 SRMMAGAFSPFYRNHNIYQISQEPYTWSSVAEASRRAMYIRYSLLPYWTIMAKASQ 749
Qy 661 GNPIARLPFPSPDPTKTYEIDISQFLIGKSIWVSPALKQGAVDADYFAGN---WEDLF 717
Db 750 GTPALRALFVEFPNDFTLADVDFQFMVGSLLVTVLEPNVEYVQGVFCDNSTWYDWY 809
Qy 718 NVSFAVGSGSKHVRUDTADHVNHVHVRGSIIVAMQGEALTTRDARKTPYQLLVASRL 777
Db 810 NHTIEV-ROYNENVTLYAPLEHINVAIRGSGVLPMQOPLTYESQNPFLNVALDRDG 868
Qy 778 NISGELFLDGENLRMGAGGGRDWTL-VKERCYVTGKSVLRSEVNVPEYASKMKWSIG 836
Db 869 SATGEYLDGVSIEUNA-----TLVSF-----TFSGVLSAV---PTGSEVVSQPLA 914
Qy 837 KVTFGVE-----NVENKTYE 853
Db 915 NVILGILTESPSSITLNGQNVSSFQ 939

RESULT 10
ID AMYG CANAL STANDARD; PRT; 946 AA.
AC O74254;
D 15-JUL-1999 (Rel. 38, Created)
D 15-JUL-1999 (Rel. 38, Last sequence update)
D 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glucoamylase 1 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)
DE (1,4-alpha-D-glucan glucohydrolase).
GN GAM1 OR GCAL.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SC5314;
RX MEDLINE=99451422; PubMed=10520161;
RA Sturtevant J., Dixon F., Wadsworth E., Latge J.-P., Zhao X.-J.,
RA Calderone R.;
RT "Identification and cloning of GCAL, a gene that encodes a cell
RL surface glucosylase from Candida albicans.";
RL Med. Mycol. 37:357-366(1999).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -!- SUBCELLULAR LOCATION: CELL WALL ASSOCIATED.
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF082188; AAC31968.1; -.
DR InterPro: IPR000322; Glyco_hydro_31.
DR Pfam: PF01055; Glyco_hydro_31.1.
DR PROSITE: PS00129; GLYCOSYL_HYDROL_F31.1; 1.
DR PROSITE: PS00707; GLYCOSYL_HYDROL_F31.2; 1.
KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW Signal; Cell wall.
FT SIGNAL 1 20 POTENTIAL
FT CHAIN 21 946 GLUCOAMYLASE 1.
FT ACT_SITE 462 462 BY SIMILARITY.
FT DOMAIN 519 532 SER/THR-RICH.
FT CARBOHYD 51 51 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 97 97 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 244 244 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 373 393 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 406 406 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 505 505 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 570 570 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 772 772 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 801 801 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 895 895 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 912 912 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 946 AA; 105804 MW; BD6B640C4BEF170 CRC64;

Query Match 29.38; Score 1393.5; DB 1; Length 946;
Best Local Similarity 34.8; Pred. No. 2.7e-86;
Matches 312; Conservative 152; Mismatches 302; Indels 131; Gaps 23;

Qy 31 ESTVVGYYVYVRSVGVDSNRQVLTKLDLKPSVYAPDIKSLNLHVSLETSLRLRI- 89
Db 56 DANAVAKYSL--VNVSLTARGLTGILKEATNIYGYDEYLNLSVEYQSDTRLANVHIE 113
Qy 90 -TDSQQORWEIPETVIPRAGNHSFRFSTEEDGNSPENNLADPSSDLVFTLHNTTTPFG 148
Db 114 PTDLTDV-FVLPBELVVK-----PKL-----EGDAKTNF--ENSDLVFE-YDEEDFG 157
Qy 149 FSVSRSSGDLIDTSPDSDSDSYTFIFKDFQLOLSSALPENSNLYGIGEHKRSRLI 208
Db 158 FEVLRSSTREVLFTSKGNP-----LVFSNQTFQNTLTPKSHS-ITGLGESIHGSLN-E 209
Qy 209 PGETMTLWNADIGSENPVNLGSHPFYMDVRGSKGNEEAGTTHGVLLNSNGMDVKYEG 268
Db 210 PGVVKTLIYANDI-ADPIDGNIYGVHPVYVYDQR-----YDNTTHGVYVWTSALQEVVVE 263
Qy 269 HRITNYVIGGVIDLYVPAGESPMVMNQYTELIGRPAMPYWSFGFHQCRGYKQVNSDLE 328
Db 264 TSLTWALSGVIDLYFFSGDPDKDVIQQYVYSEIGLPAMQPYWALGYHQCRRWGYDTVESLE 323
Qy 329 YVDCYAKAGIPLVEMWTDIDYMDCKDFDLPVNPFPEDKMQSFVDTLHNGQKQVYLID 388
Db 324 TVVENFKFDIPLLETIWSDDIDYMDGYKDFNDPYRPTDKFRKFLDLDLHNSQHYYVPIFD 383
Qy 389 -----PGIGVDSYGYTNRGMEADVPIKR-NGEPLVGEVWPGKYVFPDFLNPAAATFW 440
Db 384 AAIYVPPNPNATDDDYEPFHLGNESDVLKQPDGSLYIGAVWPGYTVFPDFLANTOEYW 443
Qy 441 SNEIKMFQELPLDGLMIDMNELSNF-----466
Db 444 NKMFQDWIERIPFDGIWTDNMNEVSSFCVSGSGDTRYFDNPVHPFPFVGYSGSDYFLGFDK 503
Qy 467 -----ITSPLSGSSSLDD-----PPYKINNSGDKRPIN 494
```



```
FT CARBOHYD 1339 1339 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 1353 1353 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 1402 1402 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 1747 1747 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 1762 1762 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT VARIANT 1097 1097 Q -> P (in disaccharide intolerance I;
FT exhibits intracellular accumulation of
FT mannose-rich SI in the Golgi).
FT CONFLICT 661 661 MISSING (IN REF. 2).
FT SEQUENCE 1826 AA; 209272 MW; 3F7E4B66DFCF9C8E CRC64;
Query Match
Best Local Similarity 38.2%; DB 1; Length 1826;
Matches 301; Conservative 138; Mismatches 258; Indels 90; Gaps 21;
QY 34 VVGYYGVYR-----SVGVDSNRQVLTKLDLKPSSVYAPDILKSLNLHVSLETSRLIR 88
DB 108 VDHGYNVQDMTTTSGVE-----AKLNRIPTSLFEGNDINSVLFTTQNTQPNRFRFK 160
QY 89 ITDSQQRWEIPEITVPRAGNHSRRFSTEEDGGNSPENNFLADPSDLVFTLH-NTTPF 147
DB 161 ITDPNRYEVPHQVY-----KEFTGPTVSDTLVDVQVAQNP 198
QY 148 GFSVRRSGDILFDTSPDSNTYFIFKQQLQSSALPENRNLYXGIEHTXRSFRL 207
DB 199 SIQVRKSNKGLFDTSGF-----LVSDQYLQISARLPD--YIYGIGSQVHKRFRH 250
QY 208 -TPGETMLWNADIGSEPDVNLGSHPPYMDVRGSKNEEGATHGVLLNSNGMDVKY 266
DB 251 DLSWKTPITFRDQPLGDNNNLYGHQTFMCI-----EDTSKSGFGVFLMNSNAMEIFI 305
QY 267 EGHRI--TYNVIIGVLDLYFAGPSPVMNVQVTELIGRPAPMPYVSGFHOCRYGKNVS 325
DB 306 QTPVTVYVTCGLDLYFLLGDTPEQVQVQQLVGLPAMPAYNWLGFQSRNWKSLD 365
QY 326 DLEVYVDGAKAGILEVMTDIDYMDGYKOTFLDPVNFPEDKMSFVDTLHNGKQYVL 385
DB 366 VYKEVVRNRREAGIFDQVTDIDYMDKKDFTYDQVAF--NGLPQFQDLHDHQKVI 423
QY 386 ILDPGIGV-----DSYGYTNGMEADYPIKEN--GEPLGVGVWQKYFFDFLNPAAAT 438
DB 424 ILDPALISIGRRANGTYATYERGNTQHVWVNESDGTPIIGSEWPLGVYDFTNPCID 483
QY 439 FWSNEIKMFOELPLDGLMDNNEISNFTSPLS--SGSSLDPPYKINNSGD--KRPIN 494
DB 484 WNAECISIFHQEVQVDGLWIDNNEVSSFLQGTGKGVNKLNYPF-----TPDILDKLMY 539
QY 495 NKTVPATSI-HFGNISEYDAHNLGLLEKAKATHQAVVDI--TKRPFILSRSTFVSSGKYT 552
DB 540 SKTICMDAVQWNG--KQYDVHSLYGSMAIAEQAVQKVPFKRSFILTRSTFAGSGRHA 597
QY 553 AHMTGDNRAKWEDLAYSIPGILNFGILGIPMYGADICGFSHDHTEELCRRTOLGAYYFP 612
DB 598 AHWLGDNTASQWMSITGMLEFSLGPIPLVGADICGVVAETTELCCRWWQLGAYYFP 657
QY 613 ARDHSLSLTARQELYLWDS--VASSARKVLGLRMELLPHELYTLMEYHVSNGNLPARLPLF 669
DB 658 SRHNSDGYEHQDPAFFGQNSLLVKSRSQYLTIYTLPLFFJYLYKAVHVGETVARVL 717
QY 670 FSPQDTKTYEIDSOFLIGKSIWSPALKQGAVADVAPFAGNWFDFLNFYFVAGDGS- 728
DB 718 HEFYEDTNSWIEDTFLWGPALLITPVLKQGAADTVSAVIPDAIWDY-----ESGA 768
QY 729 -----KHVRLDTADHNVNVHREGSIVAMQCEALTTRDARTPYQLLVASRLNISGE 782
DB 769 KFPWRQRDMYLPADKIGLHRLGGYIPIQEPDPTTASRKNPLGLIVALGENNTAKGD 828
QY 783 LFLDDGE 789
DB 829 FFWDDGE 835
```

```
RESULT 12
SUI'S RABIT
ID -SUI'S RABIT STANDARD; PRT; 1826 AA.
AC P07768;
DT 01-AUG-1988 (Rel. 08, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sucrase-isomaltase, intestinal [Contains: Sucrase (EC 3.2.1.48);
DE Isomaltase (EC 3.2.1.10)].
GN SI.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86245069; PubMed=3755079;
RA Hunziker W., Spiess M., Semenza G., Lodish H.F.;
RT "The sucrase-isomaltase complex: primary structure, membrane-
RT orientation, and evolution of a stalked, intrinsic brush border
RT protein."
RL Cell 46:227-234 (1986).
RN [2]
RP PRELIMINARY SEQUENCE OF 1-37 AND 1007-1014.
RX MEDLINE=83105704; PubMed=7152027;
RA Stoeckroem H., Noren O., Christiansen L.A., Wacker H., Spiess M.,
RA Bigler-Meier B., Rickli E.E., Semenza G.;
RT "N-terminal sequences of pig intestinal sucrase-isomaltase and pro-
RT sucrase-isomaltase. Implications for the biosynthesis and membrane
RT insertion of pro-sucrase-isomaltase."
RL FEBS Lett. 148:321-325 (1982).
CC -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE FINAL STAGE OF
CC CARBOHYDRATE DIGESTION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of sucrose and maltose by an
CC alpha-D-glucosidase-type action.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,6-alpha-D-glucosidic linkages
CC in some oligosaccharides produced from starch and glycogen by
CC alpha-amylase, and in isomaltose.
CC -!- SUBUNIT: THE RESULTING SUCRASE AND ISOMALTASE SUBUNITS STAY
CC ASSOCIATED WITH ONE ANOTHER IN A COMPLEX BY NON-COVALENT LINKAGES.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Brush border.
CC -!- PTM: THE PRECURSOR IS PROTEOLYTICALLY CLEAVED WHEN EXPOSED TO
CC PANCREATIC PROTEASES IN THE INTESTINAL LUMEN.
CC -!- PTM: N- AND O-GLYCOSYLATED.
CC -!- PTM: SULFATED (BY SIMILARITY).
CC -!- MISCELLANEOUS: THERE IS A HIGH DEGREE OF HOMOLOGY BETWEEN THE
CC ISOMALTASE AND SUCRASE PORTIONS (41 % OF AMINO ACID IDENTITY)
CC INDICATING THAT THIS PROTEIN IS EVOLVED BY PARTIAL GENE
CC DUPLICATION.
CC -!- SIMILARITY: Contains 1 P-type (trefoil) domain.
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M14046; AAA31459.1; -
InterPro; IPR000322; Glyco_hydro_31.
InterPro; IPR000519; P_trefoil.
Pfam; PF01055; Glyco_hydro_31; 2.
Pfam; PF00088; trefoil; 2.
SMART; SM00018; PD; 2.
PROSITE; PS00025; P_TREFOIL; 1.
PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 2.
PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 2.
MultiFunctional enzyme; transmembrane; Glycoprotein; Hydrolase;
Glycosidase; Repeat; Signal-anchor; Sulfation.
INIT_MET 0
CHAIN 1 1826 SUCRASE-ISOMALTASE, INTESTINAL.
FT
```





KW Hydrolase; Glycosidase; Glycoprotein; Signal.  
 FT S:GNAL 1 25 BY SIMILARITY.  
 FT CHAIN 26 985 ALPHA-GLUCOSIDASE.  
 FT ACT\_SITE 492 492 BY SIMILARITY.  
 FT CARBOHYD 126 126 N-LINKED (GLUCAC. .) (POTENTIAL).  
 FT CARBOHYD 145 145 N-LINKED (GLUCAC. .) (POTENTIAL).  
 FT CARBOHYD 220 220 N-LINKED (GLUCAC. .) (POTENTIAL).  
 FT CARBOHYD 255 255 N-LINKED (GLUCAC. .) (POTENTIAL).  
 FT CARBOHYD 349 349 N-LINKED (GLUCAC. .) (POTENTIAL).  
 FT CARBOHYD 424 424 N-LINKED (GLUCAC. .) (POTENTIAL).  
 FT CARBOHYD 508 508 N-LINKED (GLUCAC. .) (POTENTIAL).  
 FT CARBOHYD 536 536 N-LINKED (GLUCAC. .) (POTENTIAL).  
 FT CARBOHYD 539 539 N-LINKED (GLUCAC. .) (POTENTIAL).  
 FT CARBOHYD 602 602 N-LINKED (GLUCAC. .) (POTENTIAL).  
 FT CARBOHYD 624 624 N-LINKED (GLUCAC. .) (POTENTIAL).  
 FT CARBOHYD 661 661 N-LINKED (GLUCAC. .) (POTENTIAL).  
 FT CARBOHYD 835 835 N-LINKED (GLUCAC. .) (POTENTIAL).  
 FT CARBOHYD 881 881 N-LINKED (GLUCAC. .) (POTENTIAL).  
 FT CARBOHYD 929 929 N-LINKED (GLUCAC. .) (POTENTIAL).  
 FT CARBOHYD 957 957 N-LINKED (GLUCAC. .) (POTENTIAL).  
 SQ SEQUENCE 985 AA; 108704 MW; 3E9AE0AE38209E CRC64;  
 Query Match 28.6%; Score 1357; DB 1; Length 985;  
 Best Local Similarity 34.4%; Pred. No. 8.5e-84;  
 Matches 298; Conservative 142; Mismatches 290; Indels 136; Gaps 19;  
 QY 31 ESTVGVGVYVRSVGSVDNRQVLTAKLDLI-KPSSVYAPDIKSLNLHVSLETSERLRIRI 89  
 DB 62 QSVCPG----KASDVKHSQGFASLEAGPCNVYGTVDVSLTLTVEYQAKRLNLIQI 117  
 QY 90 T----DSSQORNEI-PETVIRAGNHSRPRFSTEEDGGNSPENNFADSPSDVLTFLANT 144  
 DB 118 VPTYFASNASMYLISEELVPRP-----KASQN--ASVPQSDFFVWSWSE 160  
 QY 145 TPEGFSVSRSSGDLFDTPSSDSNTYFIKDFQLQSSALPENRSLNYGIGHEHTKRS 204  
 DB 161 PSFNFKVIRKATGDVLNFT-----KGSTLVYENQFIEFVTLTPE-EYNLYGLGERMQ- 212  
 QY 205 FRLIPGETMTLWNAIDIGSENPOVNLGSHFFYMDVRGSK-----GNE 246  
 DB 213 LRLLENANLTLYAADI-ADPIDNIYGHAFYLDTRYKVGQGNKSHITVKSSEAPQSE 271  
 QY 247 EAGTHGVLLNSGMDVYEGHRTYNYVIGGVLDLYVFGSPENVMNQY-TELIIGRPA 305  
 DB 272 YVSYSHGVLRLNAHQEILLRQDKLWRTLGSGVSLTFFSGPTQAEVTKQYQLSTWGLPA 331  
 QY 306 PNEYVSFGHQCRGYGVYVSDLEYVDGVYAKAGIPLVEMWTDIDYMDGYKDFTLDPVNP 365  
 DB 332 MQQYTLGPHQCRWGNNSSEFEDVLNPFERFEIPEYLWADIDYMHGYRNFNDQHRFS 391  
 QY 366 EDKMQSFVDTLKNGQKYLILDPGI-----GVDSYGYTYNRGMEADVFIKR-NGEPLYL 418  
 DB 392 YEEGKFLKLHAGGRWRPIVDGALYIPNENASDAYETIDRGAKDDVFIKNPDGSLYI 451  
 QY 419 GEWPQKVTYPPDLNPAATFNSNEMFQETILPDGLMIDNNEISNFI----- 467  
 DB 452 GAWPQYTYTPDHPHKKASDFWANELVTVWNKLHYDGVWYDMAEVSSFCVCGTGNLSM 511  
 QY 468 -----TSP 470  
 DB 512 NPAHPFPALPGFPGNVVYDPGFTNTNATEAASAGASAASSTTTGAPYRITP 571  
 QY 471 LSSGSLDDPPYKINNSGDKRPIINKTVPATSIHFGNISSEYDAHNLVGLLEAKATHQAVV 530  
 DB 572 TPGVRNVDPHYVINVHQCHDLSVHAISPNSTHSDGVEYDVHSLYGHQGINATYHGLL 631  
 QY 531 DI-TGKRPFILSRSTVSSGKYTAHTGNAKWEIDLAYSIFGLNFGFLGPIPMVGADIC 589  
 DB 632 KVENKRPFIARSTESGSKWAGHWGDNFNSKWSMFFSISQALQFSLFGIPMGFVDTC 691  
 QY 590 GRSHDTTEELCRWIOGLGFYFPARDHSSLGTRQELYLWDSVASSARKVGLRMLPLPH 649  
 DB 692 GFNGNTDEELCNRMQLSAFFFPYRNHNVLISAIPQEPYRWASVIDATKAAMNIRIALLPY 751

QY 650 LYTLWYEAHVSGNPIARPLFFSPQDTTKTYEIDSOFLIGKSIWVSPALKOGCAVADAYFP 709  
 DB 752 FYTLFHLAHTTGSVMEALAWEPNDPSLAAGTQFLVGVSVWVIVPLEPQVDVTVGVFP 811  
 QY 710 -AGN---WFDLFNYSFAVGSQGHVRLDTPADHVNHVHREGSIVAMQGEALITTRDARKT 765  
 DB 812 GVGHGEVWYDWSQT-AVDKAPGVNTTISAPLGHIPVFEVGGSLPQHEVALITTRDARKT 870  
 QY 766 PYQLLVVASRLNISGELFLDDGENL 791  
 DB 871 PWSLTLSSNGTASGQLYLDDGESV 896  
 RESULT 14  
 ID -AMYG DEBOC STANDARD; PRT; 958 AA.  
 AC P22861; Q92336;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-AUG-1991 (Rel. 19, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Glucoamylase 1 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)  
 DE (1,4-alpha-D-glucan glucohydrolase).  
 GN GAMI.  
 OS Debaryomyces occidentalis (Yeast) (Schwanniomyces occidentalis).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Debaryomyces.  
 OX NCBI\_TaxID=27300;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX STRAIN=ATCC 26076;  
 RX MEDLINE=91071592; PubMed=1979298;  
 RA Dohmen R.J., Strasser A.W.M., Dahlems U.M., Hollenberg C.P.;  
 RT "Cloning of the Schwanniomyces occidentalis glucoamylase gene (GAMI1)  
 RT and its expression in Saccharomyces cerevisiae.";  
 RL Gene 95:111-121(1990).  
 RN [2]  
 RP SIMILARITY TO OTHER FAMILY 31 MEMBERS.  
 RX MEDLINE=92077121; PubMed=1743281;  
 RA Naim H.Y., Niermann T., Kleinhaus U., Hollenberg C.P.,  
 RA Strasser A.W.M.;  
 RT "Striking structural and functional similarities suggest that  
 RT intestinal sucrase-isomaltase, human lysosomal alpha-glucosidase and  
 RT Schwanniomyces occidentalis glucoamylase are derived from a common  
 RT ancestral gene.";  
 RL FEBS Lett. 294:109-112(1991).  
 CC -!- FUNCTION: THIS GLUCOAMYLASE HAS A SPECIFICITY TOWARD BOTH  
 CC ALPHA-1,4 AND ALPHA-1,6 LINKAGES.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-  
 CC glucose residues successively from non-reducing ends of the chains  
 CC with release of beta-D-glucose.  
 CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; M60207; AAA33923.1; -  
 DR PIR; JN0102; JN0102.  
 DR InterPro; IPR000322; Glyco\_hydro\_31.  
 DR Pfam; PF01055; Glyco\_hydro\_31\_1.  
 DR PROSITE; PS00129; GLYCOSYL\_HYDROL\_F31\_1; 1.  
 DR PROSITE; PS00707; GLYCOSYL\_HYDROL\_F31\_2; 1.  
 KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;  
 KW Signal.  
 FT SIGNAL 1 22 POTENTIAL.  
 FT CHAIN 23 958 GLUCOAMYLASE 1.  
 FT ACT\_SITE 470 470 BY SIMILARITY.  
 FT DOMAIN 26 41 SER-RICH.

FT DOMAIN 530 542 SER/THR-RICH  
 FT CARBOHYD 61 61 N-LINKED (GLNAC. . .) (POTENTIAL)  
 FT CARBOHYD 78 78 N-LINKED (GLNAC. . .) (POTENTIAL)  
 FT CARBOHYD 107 107 N-LINKED (GLNAC. . .) (POTENTIAL)  
 FT CARBOHYD 197 197 N-LINKED (GLNAC. . .) (POTENTIAL)  
 FT CARBOHYD 403 403 N-LINKED (GLNAC. . .) (POTENTIAL)  
 FT CARBOHYD 416 416 N-LINKED (GLNAC. . .) (POTENTIAL)  
 FT CARBOHYD 513 513 N-LINKED (GLNAC. . .) (POTENTIAL)  
 FT CARBOHYD 580 580 N-LINKED (GLNAC. . .) (POTENTIAL)  
 FT CARBOHYD 602 602 N-LINKED (GLNAC. . .) (POTENTIAL)  
 FT CARBOHYD 813 813 N-LINKED (GLNAC. . .) (POTENTIAL)  
 FT CARBOHYD 907 907 N-LINKED (GLNAC. . .) (POTENTIAL)  
 SQ SEQUENCE 958 AA; 106507 MW; 47938DB9BC308260 CRC64;

Query Match 28.4%; Score 1347; DB 1; Length 958;  
 Best Local Similarity 34.5%; Pred. No. 3.9e-83;  
 Matches 326; Conservative 136; Mismatches 320; Indels 164; Gaps 28;

QY 8 PNIFIVVVFSSQVLEEBESTVGVYVRSVGVDSNRQVLTAKLIDIKPSSVTA 67  
 DB 57 PNI.F-----NDSAV-----DANAAGYDL--VNTVTPRGUTGLKLEATNIYG 100  
 QY 68 PD'KSLNHLVSLTSLRLRI--TDSSQORWEIPETVIPRAGNHSPPRFSTEEOGNSP 125  
 DB 101 YDFDYLNLTVYQADRLNVHIEPTDLSDV-FVLPEHLVVKP-----LVE---GDAQ 148  
 QY 126 ENNFIADPSDLVFTLHNTTPFGFSVRSRSSGDILFDTSFSDSDSNITYFIPKDFQLSS 185  
 DB 149 SYNF-----DNCSELYFEYSN-TDFSFEVIRSTKEVLFSTKGNP-----LVFENQFIQNS 198  
 QY 186 ALPENSRLNYGICEHTKSRPLI--PGETWTLNADIGSEPNVLYGSHPFYMDVRGSK 243  
 DB 199 SLFKNHV-ITG.GESI---HGLVNEGSKVTLFANDVG-?PIDGNIYGVHPYVLDOR--- 250  
 QY 244 GNEEAGTTHGVLLNSNGMDVKYEGHRITVNYGGVIDLYVFAGPSPEMVMNQYTELIGR 303  
 DB 251 --YDTEHTHAVYRTSAIQEVLIGEESITWRALSGVIDLYFFSGTPKDAIQOYVKEIGL 308  
 QY 304 PAMPYKSGFHGCRGYKQVNSDLEVVGYAKAGIPLVMMWTDIDMGYKDFLDPVN 363  
 DB 309 PAFQPYWSLGYHOCRWGYDTIEKLSVFNFKFNPLETISDIDYMSYKDFYDPRH 368  
 QY 364 FPEDKMQSFVDTLHKNGQKQVILID-----PGIGVDSYGYTNGMEADVFIRK-NGE 415  
 DB 369 FPLDEYRKELDELHKKNHQVPIIDALVYVNPNNATDNEYPFPHGNTDVFKNPDGS 428  
 QY 416 PYLGEVMPGKYVFPDFLNPAAATFWSNEIKMFOEILPLDGLTIDMNLNFI----- 467  
 DB 429 LYICAVWQ-VTLFSRFLSRKSDM-DKVIKMDWYELTPFDGIWADNMNEVSSFCVSGGTGK 486  
 QY 468 -----TSPLSSGSSLDG-- 479  
 DB 487 YFENPAYPTVSGKATSPVGVFDVSNASEWKSIOSSISATKATSTSSVSSSSSTIDYM 546  
 QY 480 -----PPYKINNSGDKRPINNKTVPATSIHFGNISEYDAHNLGLLEAKATHQ 527  
 DB 547 NTLAPGKGNINYPPALYNMQGSDLAHAVSPNATHAGTVEYD.IHNLGYLQENATVH 606  
 QY 528 AVVDI-TGKRPFILSRSTFVSSGKYTAHWTGDNAKWEIDLAYSIPGLNFGFLGIPMVA 586  
 DB 607 ALLEVFNKRPFMISRSTFPFRAGKWTGHMGGDNTADWAYFYFIPQAFSGMGIAGLPFFGA 666  
 QY 587 DICGFSHDTTEELCRWIOLGAFYPPARDHSSLGTAQOELYLWDSVASSARKVLGLRML 646  
 DB 667 DVCGFNGNSOELCSRMWQGSFPFFFRHNYHUYGAIDQEPYVWESVAEATRTSMARYLL 726  
 QY 647 LPHLYTLMEYAHVSGNPIARPLFFSPQDTTKTYEIDSQFLIGKSIWVSPALKQGAVALDA 706  
 DB 727 LPYYTYLLHSHETITGLPILRAFQSKQPNDRSLSGVDNQFVVGDLVTVPLPBGVDKVG 786  
 QY 707 YFP-AG-----NWFDLNTYFVAGDGSQGHVRLDTPADHNVVHVREGSIVAMQGBALT 759  
 DB 787 VFFGAGKEEYVYDWTQREVFH-----KDGKNETLDAPLGHIPLHIRGGNVLPQEPGYTV 842

QY 760 RDARKTPYQLLVASRLNLSGELFDDGGENLRMGAGGGRDWTLPKRCVTKSVLVR 819  
 DB 843 AESRQNPFLGTLVALDNDGKAQSLYLDLDDGESLVSF-----SVSDN 889  
 QY 820 SEVVNPEYASMKWSIGKVTFFVG-----PENVENVKTVEVRT 856  
 DB 890 TLSASPSGDYKADQPLANVTILGVGHKPKSVKFENANVDTFYKAST 935

RESULT 15  
 ID YAJ1 SCHPO STANDARD; PRT; 993 AA.  
 AC Q09901.  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Putative family 31 glucosidase C30D11.01c precursor (EC 3.2.1.-).  
 GN SPAC30D11.01C OR SPAC56F8.01.  
 CS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 CX NCBI TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Gentles M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidaigo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skellton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,  
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lebrach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu B., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe.";  
 RL Nature 415:871-880(2002).  
 CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.eib-sib.ch/announce/>  
 CC or send an email to [license@eib-sib.ch](mailto:license@eib-sib.ch)).  
 CC -----  
 CC EMBL; Z67961; CAA91887.1; -.  
 DR EMBL; Z67928; CAA93572.1; -.  
 DR PIR; T38598; T38598.  
 DR GeneDB Spombe; SPAC30D11.01c; -.  
 DR InterPro; IPR000322; Glyco\_hydro\_31.  
 DR Pfam; PF01055; Glyco\_hydro\_31; 1.  
 DR PROSITE; PS00129; GLYCOSYL HYDROL F31.1; FALSE NEG.  
 DR PROSITE; PS00707; GLYCOSYL HYDROL F31.2; 1.  
 KW Hypothetical protein; Hydrolase; Glycosidase; Glycoprotein; Signal.

FT	SIGNAL	1	24	POTENTIAL.	
FT	CHAIN	25	993	CUTATIVE FAMILY 31 GLUCOSIDASE	
FT				C30D11.01C.	
FT	DOMAIN	24	36	POLY-SER.	
FT	CARBOHYD	7	7	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	44	44	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	89	89	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	121	121	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	138	138	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	161	161	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	169	169	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	232	232	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	361	361	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	386	386	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	393	393	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	423	423	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	447	447	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	480	480	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	488	488	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	545	545	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	548	548	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	614	614	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	673	673	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	814	814	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	826	826	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	835	835	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	846	846	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	910	910	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	940	940	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	987	987	N-LINKED (GLCNAC.	.) (POTENTIAL).
SQ	SEQUENCE	993 AA;	111043 MW;	EB9471F19AAA9BC9 CRC64;	
Query Match					28.3%; Score 1345.5; DB 1; Length 993;
Best Local Similarity					33.6%; Pred. No. 5.2e-83;
Matches 308; Conservative 151; Mismatches 302; Indels 157; Gaps 24;					
QY		38	GYVTSVGVDSNRQVLTAKDLIKPSS-VIAPDVKSLNHLVLETSELRIRITDSSQOR	96	
DB		84	GYQARNISEYS--YGVLAILELAGDACAAYGTDYPLLNLNVSYDTEERVHISISDLNQTO	141	
QY		97	WEIPETVIPRAGNHSRPFSTEEDGNSP-----ENNELADPSDLVFTLHNTTPFGSVS	152	
DB		142	FOL-----SNRRDWADPLFYRSSNFGNLQYNFSF---NTDPFEFWIT	182	
QY		153	RRSGDILFTSPDSSDSNTYFIKQDFLQLSALPENRSLNYGICEHTKRSFRILIPGET	212	
DB		183	RIACQVLFDTTRGNP-----LIFEDQYIELTNNVED-VNYVGL-SGSQSQSFRLGNNLT	234	
QY		213	MTLWADIQSENPVNLVYGHSPFYMDVR-----GSKGNEEAGTHGVLLNLSNGMDVKYEG	268	
DB		235	KTFW-ATGYSDSPKANMYGSHPFYMEQRYIPIGTT-NTYTSASHGVLMLSNGMEVLLRS	292	
QY		269	HRITVNIIGGVIDLVEYFAGP--SPENVMNQYTELIGRPAPMPYWSFGHQCRYGKXVSD	326	
DB		293	TVIKYRMIGGIIDLTVYSGSTVSPKXITQYVQSGTGTPTMQYNLSLGFQMSRWGYRTLSD	352	
QY		327	LEYVVDGYAKAGILEVMMWTDIDYMDGYKDFLDPVNPPEPDKQSFVDTLHKNGQKYYLI	386	
DB		353	LINMRSYLNASNIPTGFWNDIDYNSFRTFTVNSTAFPPNQTLDFRSLDESHQHYVYV	412	
QY		387	LDPGI-----GVDSYGTNRGHEADVFIKR-NGEPLYLGEVWPCKYVFPDFLNPAAT	438	
DB		413	LDPAIYAANPNKASDRTPYYPYSGFEDNIFIKPNPGSAVVGMAWPGVYVYVDPFTNPAVLQ	472	
QY		439	FWSNEI-----KMFQELPLDGLWIDMNLNF-----ITSPLSSGSS	476	
DB		473	YWKQILNLSTAFGNYSYDLFPGLGLDLMNEPSTFCIGSCGSDLLKLNPNVHPFSLPGD	532	
QY		477	LDO-----	479	
DB		533	VDNKYVYPEDFNATNTTEYKSVSRASQSQYKATATSEKSHETPSESILINGKPEFSINY	592	
QY		480	PPYKINNSGDKRPINNKTVPATSIHFGNISEYDAHNLYGLLEAKATHQAWVDITGK-RPF	538	

Search completed: October 27, 2003, 10:20:38

Job time : 18.5417 secs

Db	593	PPVALDTRDTETHDLAQGVSPNATMEGNTLRYNLFTNYGYSESKISPEALNSIQPNIRPF	652
QY	539	ILSRSTFVSSSKYTAHNTGDNAAKWEDELAYSIPGILNFGILPGIPMYCADICGFSHDTTEE	598
Db	653	LLSRSTFVSGRYAAHHLGDKNSQMSVSSISLITFNLGIPMYGADVCGYNGTDEE	712
QY	599	LCRRWIOQLGAFYFPFARDHSSLTGTARQELYLWDSVASSARKVLGLRMLLPLHLYTLMYEAH	658
Db	713	LCARMMALGAFLPFYRNHNSLGSIPQEPFPMASVAEASRSIAIRYSLLPYWYTLMTAS	772
QY	659	VSGNPIARPLFFFPQDTKYEIDSOPLICKSIWSPALKQGA VAVDAYFPAGN---WFD	715
Db	773	VDGTPMVRPLFFFPKQISLASVDKQFMIGTALLISPALEPNTTYIQGIIPGNDNTIWD	832
QY	716	LFNYSFAVGDSGKHVRLDTPADHVNHVHREGSIVAMQGEALTTRDARKTPYOLLVVASR	775
Db	833	WYNHS-VINHIDYENITMSAPLGTGVNIAVRGGNIIPLQOPGYTYTESRNNPYSLLTAMD	891
QY	776	LENISGELFLDDGENLRMGAG-----GGRDWTLVKFRCVYTKSVVLRSEVVNPEYASK	830
Db	892	NGFASGSLYIDDDGISMQTNSSLSVKLNSN---TITCVVSG-----TWVSSP-----	936
QY	831	MXWSIGKVTYVGFENVEN	848
Db	937	---SLANITITLGLSNPEN	951



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 ; Search time 23.5784 Seconds  
(without alignments)  
3678.964 Million cell updates/sec

Title: US-10-043-418-4

Perfect score: 4751

Sequence: 1 MSSLHWFPNIFVWVFFSL.....EVSKSLLVGKKFEVRLRLT 902

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 76:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4751	100.0	902	T48531	alpha-glucosidase
2	2830.5	59.6	903	T09143	alpha-glucosidase
3	2719.5	57.2	913	JC5463	alpha-glucosidase
4	2451	51.6	877	S63057	alpha-glucosidase
5	2120.5	44.6	915	H96709	hypothetical prote
6	1952	41.1	855	T47534	hypothetical prote
7	1515.5	31.9	952	A32609	alpha-glucosidase
8	1500.5	31.6	864	JC4624	alpha-glucosidase
9	1360.5	28.6	1827	A23945	sucrose alpha-gluc
10	1360	28.6	1827	1 U0HU	sucrose alpha-gluc
11	1357	28.6	985	JC4217	alpha-glucosidase
12	1350	28.4	995	T50267	probable family 31
13	1347	28.4	958	JN0102	glucan 1,4-alpha-g
14	1345.5	28.3	993	T38598	probable family 31
15	1313	27.6	1841	T10799	sucrose alpha-gluc
16	1271.5	26.8	1070	T19686	alpha-glucosidase
17	1261	26.5	1743	T15893	hypothetical prote
18	1143.5	24.1	919	T16693	hypothetical prote
19	1141.5	24.0	719	JC1200	alpha-glucosidase
20	1072	22.6	856	T22575	hypothetical prote
21	868	18.3	763	AH1097	alpha-glucosidase
22	864	18.2	763	AG1460	alpha-glucosidase
23	861.5	18.1	779	AE2402	alpha-glucosidase
24	833	17.5	919	T07391	probable alpha-glu
25	820	17.3	818	AC2472	alpha-glucosidase
26	819	17.2	954	S46105	glucan 1,4-alpha-g
27	802	16.9	910	T22050	hypothetical prote
28	802	16.9	924	T22044	hypothetical prote
29	745.5	15.7	941	T32449	hypothetical prote

30 713.5 15.0 693 2 H90486 alpha-glucosidase  
31 589 12.4 742 2 S11386 sucrose alpha-gluc  
32 515.5 10.9 772 2 D91195 hypothetical prote  
33 515.5 10.9 772 2 E86042 hypothetical prote  
34 514.5 10.8 731 2 D90483 alpha-xylosidase  
35 512.5 10.8 772 2 B65167 alpha-xylosidase  
36 512 10.8 772 2 A10968 probable glycosyl  
37 501 10.5 764 2 A72394 alpha-xylosidase  
38 476.5 10.0 769 2 H97033 alpha-glucosidase  
39 472 9.9 801 2 H83737 glucosidase BH0704  
40 460 9.7 773 2 A83888 hypothetical prote  
41 444.5 9.4 983 2 B87347 glycosyl hydrolase  
42 412.5 8.7 792 2 AD0104 probable glucosida  
43 412 8.7 1090 2 AG1749 glycosidase homolo  
44 409 8.6 1091 2 AF1380 glycosidase homolo  
45 399.5 8.4 1310 2 AD1380 glycosidase homolo

ALIGNMENTS

RESULT 1

T48531

alpha-glucosidase 1 - Arabidopsis thaliana

N:Alternate names: protein T22P22.110

C:Species: Arabidopsis thaliana (mouse-ear crease)

C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 21-Jul-2000

C:Accession: T48531

R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft

submitted to the Protein Sequence Database, April 2000

A:Reference number: Z24490

A:Accession: T48531

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-902 <BEV>

A:Cross-references: EMBL:AL163814

A:Experimental source: cultivar Columbia; BAC clone T22P22

C:Genetics:

A:Map position: 5

A:Introns: 78/2: 313/1; 390/1; 605/3; 747/1

A:Note: T22P22.110

C:Superfamily: lysosomal alpha-glucosidase; sucrose/isomaltase homology; trefoil homol

Query Match 100.0%; Score 4751; DB 2; Length 902;

Best Local Similarity 100.0%; Pred. No. 8.2e-317;

Matches 902; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSSLHWFPNIFVWVFFSLRSSQVLEBEESTVVGYYVRSVGVDSNRQVLTAKLDLI 60

DB 1 MSSLHWFPNIFVWVFFSLRSSQVLEBEESTVVGYYVRSVGVDSNRQVLTAKLDLI 60

QY 61 KPSSVYAPDIKSLNLHVLSLETSELRIRITDSSQORWEIPEITVPRAGNHSRRFSTEED 120

DB 61 KPSSVYAPDIKSLNLHVLSLETSELRIRITDSSQORWEIPEITVPRAGNHSRRFSTEED 120

QY 121 GGNSPENNFLADSSDLVFTLHNTTPFGSVSRSSGDIILFTDSSPSSDNTYVIFKDKF 180

DB 121 GGNSPENNFLADSSDLVFTLHNTTPFGSVSRSSGDIILFTDSSPSSDNTYVIFKDKF 180

QY 181 LQSSALPENRNLGYIGENTKGSFRLIPGETMTLWNADIGSENPDVNLGSHPFYMDVR 240

DB 181 LQSSALPENRNLGYIGENTKGSFRLIPGETMTLWNADIGSENPDVNLGSHPFYMDVR 240

QY 241 GSKNEEAGTTHGVLLNSGMDVKYEGHRTYNNVIGGVLDLYVFAGSPSEVMNQVTEL 300

DB 241 GSKNEEAGTTHGVLLNSGMDVKYEGHRTYNNVIGGVLDLYVFAGSPSEVMNQVTEL 300

QY 301 IGRPAPMPYWSFGHQCRYGKQVNSDLEYVVDGYAKAGIPLVWMTDIDYMDGYKDFDLD 360

DB 301 IGRPAPMPYWSFGHQCRYGKQVNSDLEYVVDGYAKAGIPLVWMTDIDYMDGYKDFDLD 360

QY 361 PVNFPEDKMSQFVDTLHKNGQKYVLLIDPGIGVDSSYGTYNRGMADVFIKRNGEPIYLGE 420

DB 361 PVNFPEDKMSQFVDTLHKNGQKYVLLIDPGIGVDSSYGTYNRGMADVFIKRNGEPIYLGE 420

Db 361 PVNPPEDKMQSFVDTLHKNGQKYVLLDPLGIGVDSSYGTYNRGMEADVFKXNGEPYLGE 420  
Qy 421 VWPQKVYFPFLNPAATAATFWSNEIKMFQELPLDGLWIDMNEISNFTTSPSSGSSLDLP 480  
Db 421 VWPQKVYFPFLNPAATAATFWSNEIKMFQELPLDGLWIDMNEISNFTTSPSSGSSLDLP 480  
Qy 481 PYKINNSGDKRPIINNKTVPATSIHFGNISEYDAHNLYGLLEAKATHQAVVDITGKRPFIL 540  
Db 481 PYKINNSGDKRPIINNKTVPATSIHFGNISEYDAHNLYGLLEAKATHQAVVDITGKRPFIL 540  
Qy 541 SRSTFVSSGKYTAHTWTDGNAAKWEDLAYSIPGILNFGILFGIPMWGADICGFSHDTTEELC 600  
Db 541 SRSTFVSSGKYTAHTWTDGNAAKWEDLAYSIPGILNFGILFGIPMWGADICGFSHDTTEELC 600  
Qy 601 RRTWLQGAFFPARDHSLGTARQELYLWDSVASSARKVGLRWRLPHLYTLMYEAHVS 660  
Db 601 RRTWLQGAFFPARDHSLGTARQELYLWDSVASSARKVGLRWRLPHLYTLMYEAHVS 660  
Qy 661 GNPIARPLFFSPQDTKYEIDSOFLIGKSIWSPALKQGAVADAYFPAGNMFDLFNYS 720  
Db 661 GNPIARPLFFSPQDTKYEIDSOFLIGKSIWSPALKQGAVADAYFPAGNMFDLFNYS 720  
Qy 721 FAVGDSGKHVRLDTPADHNVNHRVREGSIWAMQGEALTTRDARKTPQVLLVVASRLENIS 780  
Db 721 FAVGDSGKHVRLDTPADHNVNHRVREGSIWAMQGEALTTRDARKTPQVLLVVASRLENIS 780  
Qy 781 GELFLDGENLRMGAGGNRDWTLVKERCYVTKSVVLRSEVNVPEYASKKWSIGKVTF 840  
Db 781 GELFLDGENLRMGAGGNRDWTLVKERCYVTKSVVLRSEVNVPEYASKKWSIGKVTF 840  
Qy 841 VGFEVNVNKTVEYVTSERLSPRISLIKTIVSDNDPREFLSVEVSKLSLVGKKFEMRLR 900  
Db 841 VGFEVNVNKTVEYVTSERLSPRISLIKTIVSDNDPREFLSVEVSKLSLVGKKFEMRLR 900  
Qy 901 LT 902  
Db 901 LT 902

RESULT 2  
T09143  
alpha-glucosidase (EC 3.2.1.20) - spinach  
C:Species: Spinacia oleracea (spinach)  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: T09143  
P:Sugimoto, M.; Furui, S.; Suzuki, Y.  
Plant Mol. Biol. 33, 765-768, 1997  
A:Title: Molecular cloning and characterization of a cDNA encoding alpha-glucosidase from  
A:Reference number: 216585; MUID:97238484; PMID:9132069  
A:Accession: T09143  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-903 <SUG>  
A:Cross-references: EMBL:D86624; NID:G2081626; PIDN:BAA19924.1; PID:G2081627  
A:Experimental source: strain Dash  
C:Superfamily: lysosomal alpha-glucosidase; sucrose/isomaltase homology; trefoil homolog  
F:Keywords: glycosidase; hydrolase  
F:145-799/Domain: sucrose/isomaltase homology <SIM>

Query Match 59.6%; Score 2830.5; DB 2; Length 903;  
Best Local Similarity 59.6%; Pred. No. 2.3e-185;  
Matches 540; Conservative 138; Mismatches 205; Indels 23; Gaps 8;

Qy 8 PNIFTVVVVFFSLR---SSQVVLREESTVGVGVVRSVGVDS-NRQVLTAKLDIKPS 63  
Db 6 PSLALGILVLLQVLVAGISTSENDEPGVIGYGVKSVKVDGSGTSLRSLTALPOLVNS 65  
Qy 64 SVYAPDILKSLNLHVSLETSLRIRITDSSQORWEIPETVTPRAGNHS-----RRFST 117  
Db 66 SVYGPDIQLLSITASLESNDRLRVITDAKRRWEIPDNILHR---HQPPPPPHSSLSL 122  
Qy 118 EEDGNSPENN----FLADPSSDLVFTLHTTPGFSVSRSSGDILFDTPSPDSSDNTY 173

Db 123 YRTLLSSFTTNRKILLSHPNSDLTFTSLINTTTPGFTISRKSTHDLVFDATPDPTNPNTF 182  
Qy 174 FIFKDOFQLQSALLPENRSLNYGIGEHTRKSRFLIPGETMTLNMADIGSENPDVNLGSH 233  
Db 183 LIFDOYLHLTSSLPGTRAHIYGLGHSKPTQLAHNQTLTMRADIPSSNPVDVNLGSH 242  
Qy 234 PPMVDRSGKNEBAGTTHGVLLNSNGMDVYKSGHRIITNNVIGCVIDLIVFAGSPBMV 293  
Db 243 PPMVDRSS---PWAGSTHGVLLNSNGMDVEYGNRIITYKVIIGIIDLIVFAGSPBMV 299  
Qy 294 MNQYTELIGRPAPIPMYPSFGPHQCRYGKYNVSDLEYVVDGYAKAGIPIEVNMTDIDYMDG 353  
Db 300 VEQFTRVIGRPAPIPMYNAFGQCRYGVDYVYELQSVVAGYAKAKIPIEVNMTDIDYMDA 359  
Qy 354 YKDFLDPVNPEDKMQSFVDTLHKNGQKYVLLDPLGIGVDSSYGTYNRGMEADVFKRN 413  
Db 360 YKDFLDPVNPEDKMQSFVDTLHKNGQKYVLLDPLGIGVDSSYGTYNRGMEADVFKRN 419  
Qy 414 GEPYLGEVMPGKYVFPDFLNPAAATFWSNEIKMFQELPLDGLWIDMNEISNFTTSPSS 473  
Db 420 GKPYLGSWFGPVYFPDFLKPALTFWDEIKRFLNLLPVDGLWIDMNEISNFTTSPSS 479  
Qy 474 GSSLDPPYKINNSGDKRPIINNKTVPATSIHFGNISEYDAHNLYGLLEAKATHQAVVDIT 533  
Db 480 GSTLDNPPYKINNSGDKRPIINNKTVPATSIHFGNISEYDAHNLYGLLEAKATHQAVVDIT 539  
Qy 534 GKRPILSRSTFVSSGKYTAHTWTDGNAAKWEDLAYSIPGILNFGILFGIPMWGADICGFSH 593  
Db 540 EKRPVLSRSTFVSSGKYTAHTWTDGNAAKWEDLAYSIPGILNFGILFGIPMWGADICGFSH 599  
Qy 594 DTTEELCRRWILQGAFFPARDHSLGTARQELYLWDSVASSARKVGLRWRLPHLYTL 653  
Db 600 NITEELCRRWILQGAFFPARDHSLGTARQELYLWDSVASSARKVGLRWRLPHLYTL 659  
Qy 654 MYEAHVSNGPIARPLFFSPQDTKYEIDSOFLIGKSIWSPALKQGAVADAYFPAGN 713  
Db 660 MYEAHVSNGPIARPLFFSPQDTKYEIDSOFLIGKSIWSPALKQGAVADAYFPAGN 719  
Qy 714 FDLFNSFAVGDSGKHVRLDTPADHNVNHRVREGSIWAMQGEALTTRDARKTPQVLLVVA 773  
Db 720 FDLFNSFAVGDSGKHVRLDTPADHNVNHRVREGSIWAMQGEALTTRDARKTPQVLLVVA 779  
Qy 774 SRLNISEGLDGENLRMGAGGNRDWTLVKERCYVTKSVVLRSEVNVPEYASKKWS 833  
Db 780 SDCGSAFGELFLDQGVETVMGVNRK--WTPVKFIAASAKQTCIITSDVSGEFAVSQKW 837  
Qy 834 SIGKTVFVGFEVNVNKTVEYVTSERLSPRISLIKTIVSDNDPREFLSVEVSKLSLVGK 893  
Db 838 VIDKVTILGLRGTQKINGTYVTRGAVTRKGDKSLKSTPDRKG-EFIVAEISGLNLLGR 896  
Qy 894 KPEMLR 899  
Db 897 EFKLV 902

RESULT 3  
JC5463  
alpha-glucosidase (EC 3.2.1.20) - sugar beet  
N:Alternate names: alpha-D-glucoside glucobhydrolase  
C:Species: Beta vulgaris var. altissima (sugar beet)  
C:Date: 17-Jun-1997 #sequence\_revision 12-Sep-1997 #text\_change 20-Jun-2000  
C:Accession: JC5463; PC4330  
R:Matsumi, H.; Iwanami, S.; Ito, H.; Mori, H.; Honma, M.; Chiba, S.  
BioSci. Biotechnol. Biochem. 61, 875-880, 1997  
A:Title: Cloning and sequencing of a cDNA encoding alpha-glucosidase from sugar beet.  
A:Reference number: JC5463; MUID:97321863; PMID:9178565  
A:Accession: JC5463  
A:Molecule type: mRNA  
A:Residues: 1-913 <MATI>  
A:Cross-references: DDBJ:D89615; NID:G2217947; PIDN:BAA20343.1; PID:G2190276  
A:Experimental source: seed; cv. NK-152  
A:Accession: PC4330  
A:Molecule type: protein



```
Qy 660 SGNPIARPLFFSPQDTKTYEIDSOFLGKSIKVSIPALOGAVAVDAYPPAGNWFDFLNY 719
Db 636 TGAPIARPLFFSPQDTKTYEIDSOFLGKSIKVSIPALOGAVAVDAYPPAGNWFDFLNY 695
Qy 720 SFAYGDSGGHVRUDTPADHVNHVHREGSIVAMGEALTTTRDARKTPYQLLVASRLNI 779
Db 596 SLAVATGTGKIVRLPARADTVNHLTGTTILPQOSALTTSRARRAFHLLVLAEDGTA 755
Qy 780 SGELFDGGENLRMGAGGGRNDMTLVKFCRYVTKG--SVVLSEVNVNPEYASKMKWSIGK 837
Db 756 SGYLFDDGDSPEYGR--RSDMSVRFNFKIPNNKCAIKVKSEVENSVAQSRITLVISK 812
Qy 838 VTFVGFENVENK--TYEVTSERLSPRISLIKTUSDNDPRLS-----VEVSKLS 888
Db 813 VLMGHRSPAPKGLTVHNSAE-----VEASSAGTRYQNAAGLGVAHIGGLS 862
Qy 889 LLVGKKEFEMRLRLT 902
Db 863 LVGGEFELKVAMS 876

RESULT 5
H96709
hypothetical protein F24J5.20 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: H96709
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Pederspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tailor,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: H96709
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-915 <STO>
A:Cross-references: GB:AE005173; NID:G5734722; PIDN:AAD49987.1; GSPDB:GN00141
A:Gene: F24J5.20
A:Map position: 1

Query Match 44.6%; Score 2120.5; DB 2; Length 915;
Best Local Similarity 44.8%; Pred. No. 9e-137;
Matches 422; Conservative 166; Mismatches 277; Indels 77; Gaps 15;

Qy 2 SSLHWFNIFIVVVFSLRSSQVLEEBESTVVGYYGVVRSVGVOSNRQVLTAKLDLIX 61
Db 6 SSLAFSULLALILCFSPQSY-----KTIGKGLVSI--EESPDGGFTYQGVQK 56
Qy 62 PSSVYAPDKSLNLHVSLETSELRIRITDSSQORWEIPETVIPR-----AGNHSP 112
Db 57 KNKIYSGDITTLRLVFKHETDLSRLVRHITDAKQORWEVPYNLAPREQPPQGVKIGSRK 116
Qy 113 RRFSTEEDGGSNNFNFLADPSSDLVFTLHNTTTPFGSVSRSSGDILFTDTPSSDSNT 172
Db 117 SPITVQETSG-----SELIFS-YTTDPTFAVKRSNHNELFTNT-----S 156
Qy 173 YFIFKDFLOLSSALPENRSLNYLGEHTK-RSEFLTEGETMTLWNADIGSENPDVNLVG 231
Db 157 SLVFKDQVLEISLTPKEAS-LYLGENSQANGIKLVNPEYTLTYEDVSAINLNTDLYG 215
Qy 232 SHPPFVDRGSGKNEEACTTGVLNLSNGMDVKYEGHRITVYNGVIDLVFPAGSPSE 291
Db 216 SHPMYMDLNVGKAYE---HAVLLNLSNGMDVYRGDSLTYKVIGGVDFYFIAGSPFL 272
Qy 292 MVNQYQTELIGRPAMPYVSGFQHCRQYGYKRVSDLEYVVDGYAKAGIPIEVMMWTDIDYM 351
```

```
Db 273 NVVOYQYTLIGRPAMPYVWSLGFQCRWGHNTLSVVEDVDVNYKKAKLPLDVIWDDHMM 332
Qy 352 DGYKDFLTDPVNFYFEDKMQSFVDTLHKNGQKYVILDPGIGVDSSYGTYNRGMADVFIK 411
Db 333 DGHKDFLTLPVAVYPRAKLLAFDLKIHKIMKYIIVNDPGIGVNASYGTFFQMAADVFIK 392
Qy 412 RNBEPYLGVEWPGKVPYPPDFLNPAAATFWSNEIKMFOEILPLDGLWIDMELSNFITS-- 469
Db 393 YEGKPFLLAQVWPGVPYPPDFLNPXTVSNWGDIEIKRFHDLVPIDGLWIDMELSNFSCSLC 452
Qy 470 -----PLSSG-----SSLDPPYKINNKGDKRPINNKTPVATSIHFGN 507
Db 453 TIPECKQCPSEGGPGWCCLDCKNITKTRWDDPPYKINATGVAVPGFKIATSIATHNG 512
Qy 508 ISEYDAHNLYLLEAKATHQAQVVDITGKRPFILSRSTFVSSGKYTAHTWTDGNAKAWEDLA 567
Db 513 VREYDAHSIYGFSEIATHTKGLLVQGRPPILSRSTFVSGQYAAAHWTGDNQGTWQSLQ 572
Qy 568 YSIPGILNFGLPYPMGADICGFSHDTTELCRMIOLGAFYPPFARDHSSLSGTAROEY 627
Db 573 VSIITMLNFGIFGPMVGSIDCGFPQPTTELCNRWIEVGAFYFSDRHANYSPRELY 632
Qy 628 LWDVSSARKVLGLRMLLPHLYTLMYEARVSGNPIARPLFFSPQDTKTYEIDSOFLI 687
Db 633 QWDTVADSGARNALGWRKYKILPFLVTLNYEARMTGAPIARPLFFSPFPEYCEYCNRSQFL 692
Qy 698 GKSINVSALPKQAVAVDAYFPAGNWFDFLNYSAVGDGSKHYRLDTPADHVNHVHREG 747
Db 693 GSSFMISVLEQGTVEALFPPGSHYHMFDMQTAQVSKNGKRVTLPAFLNFVNVHLYQN 752
Qy 748 SIVAMQGGALTTTRDARKTPYQLLVV--ASRLEN--ISGELFDGGENLRMGAGGGRNDWL 804
Db 753 TILPTQOGLSLSKOARTTFFSLVIAFPAGASEGATGKLYLDEDELPEMKLNGQS--TY 810
Qy 805 VKFRCYVTGKSVLVRSEVNVNPEYASKMKWSIGKTVFGFENVENVKYVEVTSERLSRSP 864
Db 811 VDFYASVNGTGMKMSQVKEGKFALSKGVWIEKVSVLGRGAGQVSEIQINGSPMTKKIE 870
Qy 865 IS-----LIKTVDNDPRLSVEVSKLSLLVGVKKFEMRLRL 901
Db 871 VSSKEHTYVIGLEDEENKSNVMEVVRGLEMLVGVKDFNMSWKN 912

RESULT 6
T47534
hypothetical protein F1612.150 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000
C:Accession: T47534
R:Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unseid, M.; Mewes, H.W.; Rudd, S.
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24468
A:Accession: T47534
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-855 <JOR>
A:Cross-references: EMBL:AL162459
A:Experimental source: cultivar Columbia; BAC clone F1612
C:Genetics:
A:Map position: 3
A:Introns: 69/3; 291/1; 427/3; 699/1
A:Note: F1612.150
C:Superfamily: lysosomal alpha-glucosidase; sucrase/isomaltase homology; trefol homo

Query Match 41.1%; Score 1952; DB 2; Length 855;
Best Local Similarity 43.8%; Pred. No. 2.7e-125;
Matches 399; Conservative 144; Mismatches 263; Indels 106; Gaps 16;

Qy 13 VVVFPSLRSSQVLEEBESTVVGYYGVVRSVGVDSNRQVLTAKLDLIKPSVVVAPDKS 72
Db 11 IILCFSSLQ-----SNAIGKYLISMEKSPDGGSGFYGLVQVQSKNYGSDITI 61
```





Query Match 31.9%; Score 1515.5; DB 1; Length 952;  
Best Local Similarity 38.1%; Pred. No. 2.4e-95;  
Matches 336; Conservative 137; Mismatches 302; Indels 107; Gaps 23;  
QY 46 VDSNRQVLTAKLRLIKPSVYAPDKSLNHLVSLTSELRIRITDSSQORWEIP-ETVI 104  
DB 141 LSSSEMGTALTTRTP-TPFKDLILRLDVMETENRHLFTIKDPAARYEVLPT-- 197  
QY 105 PRAGNHSPPRESTEDGNSPENNFLADPPSDLVFTLHNTTTPFGFSVRSRSGDILFDTS 164  
DB 198 PHVHSRAP-----SPLYSVEFSEEPFGVIVRRQDGRVLLNTT 235  
QY 165 PDSSDSNTYIFKQDFLOLSSALPNRSNLXGIGEHKTRSPRLIPGETMTLWADIGSEN 224  
DB 236 V-----APLFAQDFLOLSTSLFS--OYITGLAEHLSPMLSTSWTRITLWNRDL-APT 286  
QY 225 PDVNLGSHPPYMDVRGSKNGEAGTTHGVLLNSNGMDVKYE-GHRITYNVIGGVIDLY 283  
DB 287 PGANLYGSHPEYLA-----EDGSAHGVELLNSNAMDVVLPSPALSWRSTGGILDVY 340  
QY 284 VFAGSPSPMNNQYITELIGRAPMYPMSFGHQCRYGKXVSDLEYVDGYAKAGIPLV 343  
DB 341 IFLOGEPKSVVQQQLDVGYPFMPYMGGLGPHLCRWGYSATIRQVVENKTRAHPELDV 420  
QY 344 MWTIDIDYMGYKDFLDPVNPEDKQOSFVDTLHKNGQKYLIDPGI---GVDSYGY 400  
DB 401 QWNLDDYMSRDRDFTFNKQGRD--FPAMVQELHQGRRYMWIVDPAISGGPAGSVRYP 458  
QY 401 NRGMEADVFI-KRNGEPYLGWPGKGVYFPDFLPNAPAAATFMSNEIKMFOELPLDGLWID 459  
DB 459 DEGLRRGVFITNETGQPLIGKVMFGSTAFDPFTNPTALAWMEDVVAEFHQVQVDFGHWID 518  
QY 460 XNELSNFTITPLSS--GSSLDPPYKINNCDKRPINNKTVPATSIHFCNISEYDAHNL 517  
DB 519 MNEFSNFIREDGCPNNELENPPYVGVGG--TLQAATICASSHQFLS-THYNLHNL 575  
QY 518 GLLEAKATHQAVVDITGKRPPIILSRSTFVSSGKYTAHTWTDNAAKWEDLAYSPGILNKG 577  
DB 576 GLTEALASHRALVKARGTRPFVISRSTFAGHGRVAGHTGDMWSSWQLASSVPEILQFN 635  
QY 578 LFGIPMGADICGSHDTHTELCRRWTQJGAFYFPADHSLSLGTARQELYLWDSVASSA- 636  
DB 636 LLGVPLGADYCGFELGNTSEELCVRTQJGAFYFPFMRNHNLSLLPOEPYSFSEPAQAM 695  
QY 637 RKVLGLRMLLPHLYTLMYEARVSNPTARPLTFSPQDTKYEIDSQFLIGKSIWSPA 696  
DB 696 RKALTLRVALPHLYTLFHOAHVAGETVARPLFLEFPADSSWTVDHQLLWGEALLITPV 755  
QY 697 LKQGAVAVDAYFPAGNMFDFLNFYSF-AVGG-----DSGKHVRLDTPADHVN 741  
DB 756 LQAGKAEVTGYFPLGTWYDLQTVPIEALGSLUPPPAPPAIHSBGQWVTLPAFLDTIN 815  
QY 742 VHVREGSVAMOGAALTRDARKTPYQLLVVASRLNISELFLDGCENLBMGGG--- 798  
DB 816 VHLRAGYIIPQGGCLTITTESRQOPMALAVALTGKGGEARGFLFWDGSESLEVLERGAYTQ 875  
QY 799 ----NRDWTLVKFCYVTGKSWLRSEVNVNPEYASKKWSIGKTVFGFENV-ENVKIYE 853  
DB 876 VIFLARNTIYNELVRVTSEGAGLQ-----LQKTVLGVATAPQOVLNNG 920  
QY 854 VRTSERLSPRISLIKTVDNDPRFVSVEVSKLSLLVGKXP 895  
DB 921 VPFVSNFTYSP-----DTKVLDDICV---SLLMGEQF 947

RESULT 8  
JC4624  
alpha-glucosidase (EC 3.2.1.20) - Rhizomucor circinelloides f. circinelloides  
C:Species: Rhizomucor circinelloides f. circinelloides  
C:Date: 10-May-1996 #sequence\_revision 19-Jul-1996 #text\_change 20-Jun-2000  
C:Accession: JC4624; PC4149  
R: Sugimoto, M.; Suzuki, Y.

J. Biochem. 119, 500-505, 1996  
A:Title: Molecular cloning, sequencing, and expression of a cDNA encoding alpha-glucosidase  
A:Reference number: JC4624; MUID:96271012; PMID:8830045  
A:Accession: JC4624  
A:Molecule type: mRNA  
A:Residues: 1-864 <SUG>  
A:Cross-references: DDBJ:D67034; NID:gl498134; PIDN:BAAL1053.1; PID:gl498135  
A:Accession: PC4149  
A:Molecule type: protein  
A:Residues: 203-214; 492-495; 612-623; 715-731; 742-751; 769-778 <SUG>  
A:Note: The source is designated as Mucor javanicus IF04570  
C:Comment: This enzyme is an exo-carbohydrase, and catalyzes the splitting of an alpha well as soluble starch.  
C:Superfamily: lysosomal alpha-glucosidase; sucrose/isomaltase homology; trefoil homo  
C:Keywords: glycoprotein; glycosidase; hydrolase  
F:122-805/Domain: sucrose/isomaltase homology <SIM>  
F:187,364,466,506,568,734/Binding site: carbohydrate (Asn) (covalent) #status pre-  
F:430/Active site: Asp #status predicted

Query Match 31.6%; Score 1500.5; DB 2; Length 864;  
Best Local Similarity 38.8%; Pred. No. 2.2e-94;  
Matches 307; Conservative 140; Mismatches 252; Indels 93; Gaps 18;  
QY 66 YAPDIKSLNHLVSLTSELRIRITDSSQORWEIPETVIPRAGNHSPPRSTEDGGNSP 125  
DB 67 FCKTIKIDTVDEYETEERLHVKISDKAKQYLPDSPL-----GFERP 110  
QY 126 E-NFLADPPSDLVFTLHNTTTPFGFSVRSRSGDILFDTSPPSDSNTYIFKQDFLOLS 184  
DB 111 QIKHVSYPKSNLDPQ-YTAKPFKVKVRKDKTIFDT-----TNMPLVFEDQYLELS 163  
QY 185 SALPENRNLXGIGHTKRSFRLIPGETMTLWADIGSENPD---VNLXGSHPPFMDVRG 241  
DB 164 TKVPED-ANVIGIGVT-APPRTHNT-TLWARD----NPDDFYRNIYGAHPYQEVVD 216  
QY 242 SKGNBEAGTTHGVLLNSNGMDVKYEGHRTYINVIYGVVIDYVFAGPS--PEMWNQYTE 299  
DB 217 -----GKAHGALLMNAHGMVDVITTEGRITKYKVGILDFYFAPKSGKNDLSIAYTD 269  
QY 300 LIGRAPMYPMSFGHQCRYGKXVSDLEYVDGYAKAGIPLVWMTDIDYMGYKDFL 359  
DB 270 LIGKPMPSHMLGHHHCYGYNDKVTWKRYKEANIPLQTVWVIDIMEETKDTFF 329  
QY 360 DPVNPEDKQOSFVDTLHKNGQKYLIDPGIYGVDSYGYTNRGMEADVFIKR-NGEPL 418  
DB 330 DAVNFPQDMIGLQGLHKGQYVWVDPAISANTYEVYRGTEMVDWIKNADGSDFI 389  
QY 419 GSWPFGKVPDFLPNAPAAATFWSNEIKMFOELPLDGLWIDMNLNSNFTITPLSSG---- 474  
DB 390 GSWPFGFTFPDWMHPNATKYWNKEIIDFVDMGLVDGLWIDMNEPASFCLSGCSGKVDA 449  
QY 475 -----SSLDPPP-----YKINSGDKRPPINNKTV 499  
DB 450 GNQPRWYTYTEEQAAHTRWEKELKANGNPPEERNLLYKYALINNGAGN--LSEFTVA 507  
QY 500 ATSIHFGNISEYDAHNLGLLEAKATHQAVVDITGK-RPFIILSRSTFVSSGKYTAHTMGD 558  
DB 508 TTAHLYGNIPHYDIIHNLGHAESHITRQALIKHKNKIRPFVLTRSSFPGSGKSVGHMTGD 567  
QY 559 NAAKVEDLAYSPGILNFGILPGIPMGVADICGFSHDITTEELCRRWIQLGAPYPARDHSS 618  
DB 568 NHSFPYPLKNSIANTLNFQMFVSGYSGADVCGFNSDTTEELCTRMEIGAPYFARNHN 627  
QY 619 LGTARQELYLWDSVASSARKVLGLRMLLPHLYTLMYEARVSNPTARPLFPSPQDPTKT 678  
DB 628 NAAKQEPYLWESTABASRIANTRYEMLPFYTLFEESNRLGLGVWRPLLFEPYPAYEEL 687  
QY 679 YEIDSQFLIGKSIWSPALKQCAVAVDAYFPAGNMFDFLNFYSFVAGDGGKHVR---LDT 735  
DB 688 VSNVDQTLVGSDILLSPVLDEGKTSVKAFQPGQGWYDMYTHETLVDNKSNKKVKTTLDA 747  
QY 736 PADHVNHVREGSVAMOGAALTRDARKTPYQLLVVASRLNISELFLDGCENLBMGA 795

Db 748 PLTHPIHTRGCAIIPKTPKTYVGTFTFPYVNLVIALDKQAGRLYIDGESLEVK 807

QY 796 GCGN-----RDWT 803

Db 808 SSGYHFLQEW 819

RESULT 9

A23945

sucrose alpha-glucosidase (EC 3.2.1.48) / oligo-1,6-glucosidase (EC 3.2.1.10) - rabbit

N:Alternate names: small intestinal sucrase/isomaltase (SI)

C:Species: Oryctolagus cuniculus (domestic rabbit)

C>Date: 03-Mar-1994 #sequence\_revision 03-Mar-1994 #text\_change 24-May-1996

C:Accession: A23945; B25987; A29163

R:Hunziker, W.; Spiess, M.; Semenza, G.; Lodish, H.F.

Cell 46, 227-234, 1986

A:Title: The sucrase-isomaltase complex: primary structure, membrane-orientation, and evolution

A:Reference number: A23945; MUID:186245068; PMID:3755079

A:Accession: A23945

A:Molecule type: mRNA

A:Residues: 1-1827 <HUN>

R:Sjoestrom, H.; Noren, O.; Christiansen, L.A.; Wacker, H.; Spiess, M.; Bigler-Meier, E

FEBS Lett 148, 321-325, 1982

A:Title: N-terminal sequences of pig intestinal sucrase-isomaltase and pro-sucrase-isomaltase

A:Reference number: A25987; MUID:83105704; PMID:7152027

A:Accession: B25987

A:Molecule type: protein

A:Residues: 2-32, 'XXX', 36-38, 1008, 'N', 1010-1014, 'E' <SJ2>

R:Frank, G.; Brunner, J.; Jauser, H.; Wacker, H.; Semenza, G.; Zuber, H.

FEBS Lett 96, 183-188, 1978

A:Title: The hydrophobic anchor of small-intestinal sucrase-isomaltase. N-terminal sequence

A:Reference number: A29163; MUID:79086207; PMID:729784

A:Accession: A29163

A:Molecule type: protein

A:Residues: 2, 'VNA', 6-32, 'XXX', 36-38 <FRA>

C:Comment: Carbohydrate analysis of the mature enzyme complex indicates both N- and O-linked

C:Complex: the two product chains remain associated after cleavage

C:Function: <ISM>

A:Description: oligo-1,6-glucosidase catalyzes the hydrolysis of 1,6-alpha-D-glucosidic

A:Pathway: carbohydrate digestion

C:Function: <SUC>

A:Description: sucrose alpha-glucosidase catalyzes the hydrolysis of the 1,2-alpha-D-glucopyranosidic

A:Pathway: carbohydrate digestion

C:Superfamily: sucrase/isomaltase

C:Keywords: carbohydrate digestion; duplication; glycoprotein; glycosidase; hydrolase; m

F:2-1007/Product: oligo-1,6-glucosidase (isomaltase chain) #status experimental <ISO>

F:13-32/Domain: membrane associated #status predicted <TM>

F:43-65/Region: serine/threonine-rich

F:63-109/Domain: trefoil homology <TRF1>

F:189-840/Domain: sucrase/isomaltase homology <SIM>

F:931-977/Domain: trefoil homology <TRF2>

F:1008-1827/Product: sucrose alpha-glucosidase (sucrase chain) #status experimental <SUC>

F:1062-1734/Domain: sucrase/isomaltase homology <SIM>

F:12/Binding site: carbohydrate (Thr) (covalent) #status experimental

F:99,455,859,896,904,1235,1303,1325,1340,1354,1403,1535,1572,1748,1763,1799/Binding site: 1007-1008/Cleavage site: Arg-116 (trypsin) #status predicted

Query Match 28.6%; Score 1360.5; DB 1; Length 1827;

Best Local Similarity 35.3%; P-red. No. 2.9e-84;

Matches 312; Conservative 173; Mismatches 318; Indels 81; Gaps 24;

QY 34 VVGYYGVVRSVGVDSNRQVLTAKLDLIKPSVYAPDIKSLNLHVSLETSEIRIRITDSS 93

Db 109 VDNHGYNVE-GMTTSTGLEARLNKSTPTLFGNDINNVLTTESQTANRLRFLKIDPN 166

QY 94 QQRWEIPETVIPAGNHSPPRSTEDGNSPENNFLADPSSDLVFTLHNT-TPGFSVS 152

Db 167 NKRYEV-----PHQFVTE-----FAGPAATETLYDQVVTENPFSIKVI 204

QY 153 RRSNGDILFDTSDDSDSNTYFIFFKDFQLQSLAPENKSNLYGIGETKRSFRL-IPGE 211

Db 205 RKSNKRLFDSSIGP-----LVYSQYLIQISLRUPS--EYMYGFGHVKRFRHLYWK 256

212 TMTLNADIGSENPDVNLGSHPFYMDVRGSKGNEEAGTTHGVLLNSNGMDVYEGHRI 271

Db 257 TWPIFTRDQHTDDNNNLYGHQTFMCI-----EDTTGKSGFVFLMNSNAMEIFIQTPFI 311

QY 272 -TYNVIGSVIDLVPAGSPBMVNNQYTELIGRPAPMPYMSFGHQYGYKYNVSDLEYV 330

Db 312 VYRVIGGILDYIFLGDTPEDVYVQYQELIGRPAMPAYMSLGFQLSRWYNSLDVVKEV 371

QY 331 VDGAKAGIPLVNMWTDIDYMGYKDFTLDPVNFPEKMGQSFVDTLHKNQYKYLIDPG 390

Db 372 VRNRREALIPDPTQVSDIDYEDKKDFTYDRVAY--NGLPDFVODLHDHGGKYVILLDPA 429

QY 391 IGVD-----SSYGTYNRGMADYFIKRG--EPYLGEVWPKYVFPDFLPAAATFWSNE 443

Db 430 ISINRRASGEAYESYDRGNAQVWVNESDGTTPVIGVWPCDTPVDFTPSNCIEWANE 489

QY 444 IKMQEILPLDGLMIDNVELSNFI--TSPSSSGSLDDPPYKINNKGDKRPNNKTVPAT 501

Db 490 CNIFHOEVNYDGLMIDMNEVSSVFGSGNKGCDNTLNPYP-IPDIVDKL-MYSKTLCMD 547

QY 502 SIHFGNISEYDAHNLYGLLEAKATHQAVVDI--TGKRPPIILSRSTFVSSGKYTAHMTGDNA 560

Db 548 SVQFWG-KQYDVHSLYGYSAIATERAVERVFPNKRSPILTRSTFAGSGRHAHWLGNT 606

QY 561 AKWEDLAYSPILNFGIFGIPMYGADICGFSHDTTBELCRRWITQLGAFYFPFARDHSLG 620

Db 607 ATWEQMEWSITGMLEFGLGMPLVGADICGFAETTELCCRWMQLGAFYFSPSRHNADG 666

QY 621 TARQE--LYLWDS--VASSAKYVLGLRMLPLHLYTLMEAHVSGNPIDARLPFSFPDOK 677

Db 667 FEHQDPAFFQGSLLVKSRRHYLNIRYTLPLFLYKAHAPGETVARPVLHFEFYEDTN 726

QY 678 TYEIDSOFLIGKSIMVSPALKQGAVDAYFPAGNWFDFNFYSAVAGD---SGKHVRLD 734

Db 727 SWEDREFLWGPALLITPVLTOGAETVSAYIPDAVWD-----YETGAKPRKQRVEMS 781

QY 735 TPADHVNHVREGSIVAMQGEALTRDARKTPYQLLVASRLNISELGFLDDGLENLRMG 794

Db 782 LPADKIGLHLRGVYIIPQAPVTTTASRMPLGLIILALDDNTAVGDFFWDDGETKDTV 841

QY 795 AGGNRRDWTLKFCYCYTGKSVLVRSVNVPEVASKKNIGKVTFFGVFPENVENKTYEV 854

Db 842 Q---NDRIYILYFAVSNLNLITCHEL-----YSEG--TTLAFTQIKILGVET 886

QY 855 RTSERLSPRISLIKTYDNDPRFLSVEYSKLSLLVGKKFEMR 898

Db 887 VTQVTVANNQSMSTHSNFTYDPSNQVLLIENLNFNLGRNFRVQ 930

RESULT 10

UUHJ

sucrose alpha-glucosidase (EC 3.2.1.48) / oligo-1,6-glucosidase (EC 3.2.1.10) (validat

N:Alternate names: limit dextrinase; small intestinal sucrase/isomaltase (SI)

C:Species: Homo sapiens (man)

C>Date: 19-Nov-1988 #sequence\_revision 24-May-1996 #text\_change 08-Dec-2000

C:Accession: S36082; A27326; S24329; A61136

R:Lacasa, M.

submitted to the EMBL Data Library, December 1991

A:Reference number: S36082

A:Accession: S36082

A:Molecule type: mRNA

A:Residues: 1-1827 <LAC>

A:Cross-references: EMBL:X63597; NID:g36644; PIDN:CAA45140.1; PID:g36645

R:Green, F.; Edwards, Y.; Hauri, H.P.; Povey, S.; Ho, M.W.; Pinto, M.; Swallow, D.

Gene 57, 101-110, 1987

A:Title: Isolation of a cDNA probe for a human jejunal brush-border hydrolase, sucrase

A:Reference number: A27326; MUID:88112852; PMID:2962903

A:Accession: A27326

A:Molecule type: mRNA

A:Residues: 1-661, 'X', 663-678 <GRE>

A:Cross-references: GB:M22616

R:Chantret, I.; Lacasa, M.; Chevalier, G.; Ruf, J.; Islam, I.; Mantei, N.; Edwards, Y.

Biochem. J. 285, 915-923, 1992

A;Title: Sequence of the complete cDNA and the 5' structure of the human sucrase-isomaltase  
A;Reference number: S24329; MUID:92359963; PMID:1353958  
A;Accession: S24329  
A;Molecule type: mRNA  
A;Residues: 1-661, 'F', 663-931 <CHA>  
A;Cross-references: EMBL:X63597  
R;Gorvel, J.P.; Ferrero, A.; Chambraud, L.; Rigal, A.; Bonicel, J.; Maroux, S.  
Gastroenterology 101, 618-625, 1991  
A;Title: Expression of sucrase-isomaltase and dipeptidylpeptidase IV in human small intestine  
A;Reference number: A61136; MUID:91317403; PMID:1677636  
A;Accession: A61136  
A;Molecule type: protein  
A;Residues: 2-14, 'F', 16-20; 1008-1015, 'E', 1017-1021, 'TX', 1024 <GOR>  
C;Genetics:  
A;Gene: GDB:S1  
A;Cross-references: GDB:120377; OMIM:222900  
A;Map position: 3q25.2-3q26.2  
C;Complex: the two product chains remain associated after cleavage  
C;Function: <ISM>  
A;Description: oligo-1,6-glucosidase catalyzes the hydrolysis of 1,6-alpha-D-glucosidic  
A;Pathway: carbohydrate digestion  
C;Function: <SUC>  
A;Description: sucrose alpha-glucosidase catalyzes the hydrolysis of the 1,2-alpha-D-gluc  
A;Pathway: carbohydrate digestion  
C;Superfamily: sucrase/isomaltase; sucrase/isomaltase homology; trefoil homology  
C;Keywords: carbohydrate digestion; duplication; glycoprotein; glycosidase; hydrolase; m  
F;2-1007/Product: oligo-1,6-glucosidase (isomaltase chain) #status experimental <XAT1>  
F;13-32/Domain: membrane associated #status predicted <TMM>  
F;42-60/Region: serine/threonine-rich  
F;63-109/Domain: trefoil homology <TRF1>  
F;189-840/Domain: sucrase/isomaltase homology <SIM>  
F;931-977/Domain: trefoil homology <TRF2>  
F;1008-1827/Product: sucrose alpha-glucosidase (sucrase chain) #status experimental <XAT  
F;1062-1734/Domain: sucrase/isomaltase homology <SIM2>  
F;12/Binding site: carbohydrate (Ser) (covalent) #status absent  
F;99,437,455,823,855,904,926,1235,1303,1340,1354,1403,1535,1572,1675,1748,1763,18-5/Bind  
F;1007-1008/Cleavage site: Arg-Ile (trypsin) #status predicted

Query Match 28.6%; Score 1360; DB 1; Length 1827;  
Best Local Similarity 38.2%; Pred. No. 3.1e-84;  
Matches 301; Conservative 138; Mismatches 258; Indels 90; Gaps 21;

QY 34 VVGYYGVVVR-----SVGVDSNRQVLTAKLDLKPSSVYAPDIKSLNLHVSLETSELRIR 88  
DB 109 VDNHGVNQDMTTSIGVE-----AKLNRPSTLFGNDINSVLTQTQNPFRFK 161  
QY 89 ITSSQQRWIPETVIPRAGNHSPPRSTEDGNSPENNFLADPSSDLVFTLH-NTTFF 147  
DB 162 ITDPNNRRYVPHQV-----KEFTGPTVSDTLTYDKVQAQNP 199  
QY 148 GFSVSRSSGDLFDTSDDSDNTYFIKQDFQLSSALPENRSLYGIQHTKRSFL 207  
DB 200 SIQVIRKSNKTLFDTSIGP-----IVYSDYLQISARLPSPD--YYIGGEQVHKRFRH 251  
QY 208 -IPGCTMTLWNADIGSENPDVNLGSHPPFYMDVRGSKGNEEAGTHGVLLNSNGMDVKY 266  
DB 252 DLSWKTPFIPTDQLPGDNNNLYGHOTFFMCI-----EDTSKGSFGVFLMNSNAMEIFI 306  
QY 267 EGHRI-TYVNLGGVIDLYVAGSPENVMNQYTELICRPAPMFWSGFFHQCYGYKNVS 325  
DB 307 QPTPIVTRYVTGGILDFYLLGDTPEQVQYQQLVGLPAMPAYMNLGFLSRWNYKSLD 366  
QY 326 DLEYVVDGYKAGIPLEVMWTDIDYMGYKDFLTLDVNFPPEDKMQSPVDTLHNGKQKYL 385  
DB 367 VKKEVVRNREAGLPFTQVTDIDYMDKKDFTYDQVAF--NGLPQVQLDHDHGQKVI 424  
QY 386 ILDPGIGV-----DSSYGYTNRMGEADVFIKRN--GEFYLGEVMPGKVPYFPDFLNPAAT 438  
DB 425 ILDPAISIGRRANGTTVATVERGNTQHWINESDGSPTIIGVWPGLVTVYDFTFNPNCID 484  
QY 439 FWSNEIKWFQELLPLDGLWIDMELSNFTISPLS--SGSSLDPPPYKINNSGD--KRPIN 494  
DB 485 WMANECSIFHQEQVDGLWIDMNEVSSFIQSGTKGCNVNKLNTYPPF-----TPDILDKLKY 540

QY 495 NKTVPATSI-HFGNISEYDAINLYCLLEAKATHOAVVDI-TGKRPPFILSRSTFTVSSGKYT 552  
DB 541 SKTICMDAVQNWG--KOYDVHSLYGYSMIAITEAQVQVFPNKRFSILTRSTFAGSGRHA 598  
QY 553 AHWTDNAAKVEDLAYSPGILNFGILPGIPMVGADICGFSHDTTTEELCRRTQLGAFYFF 612  
DB 599 AHWLGDTASHEQWESITGMLESLFGIPLVGADICGFVAETTEELCRRMWQJGAFYFF 658  
QY 613 ARDHSSSLGTARQELYLWDS---VASSARFKVLGLRMRLPHLYTTLMYEAHVSGNPIARPLF 669  
DB 659 SRNHSNDCGYEHQDPAPFQNSLLVKSSRQYLIRVTLPLPFLTYLTKAHVFEGETVARPVL 718  
QY 670 FSPFQDTTYEIDSOFLIGKISIMVSPALKQCAVADYAFYFAGNWFEDLFNYSFAVGDSG- 728  
DB 719 HEFYEDTNSWIEDTEFLMGPAALLITPVLKQAGDTSAYIPDAIWDY-----ESGA 769  
QY 729 -----KHVRLDTPADHNVHVREGSIVAMQGEALTRDARKTPVQLLVASRLNISGE 782  
DB 770 KRPWRKQVDMYLPADKLGILHGGYIPIQEPDVTITASRKNPLGLIVALGENTAKGD 829  
QY 783 LFLDDGE 789  
DB 830 FFWDDGE 836

RESULT 11  
JC4217  
alpha-glucosidase (EC 3.2.1.20) - Aspergillus oryzae  
C:Species: Aspergillus oryzae  
C:Date: 21-Nov-1995 #sequence\_revision 08-Feb-1996 #text\_change 21-Jul-2000  
C:Accession: JC4217  
R;Minetoki, T.; Gomi, K.; Kitamoto, K.; Kumagai, C.; Tamura, G.  
Biosci. Biotechnol. Biochem. 59, 1516-1521, 1995  
A;Title: Nucleotide sequence and expression of alpha-glucosidase-encoding gene (agda)  
A;Reference number: JC4217; MUID:96032211; PMID:7549103  
A;Accession: JC4217  
A;Molecule type: DNA  
A;Residues: 1-985 <MIN>  
A;Cross-references: DDBJ:D45179; NID:gl054564; PIDN:BAA08125.1; PID:gl054565  
C;Comment: This enzyme catalyzes the liberation of alpha-glucose from the non-reducin  
portant enzyme in the food industry as the isomalto-oligosaccharides.  
C;Genetics:  
A;Gene: agda  
A;Introns: 234/2; 371/2; 428/2  
C;Superfamily: Schwannomyces glucan 1,4-alpha-glucosidase GAM1; sucrase/isomaltase h  
C;Keywords: glycoprotein; glycosidase; hydrolase  
F;152-898/Domain: sucrase/isomaltase homology <SIM>  
F;126,145,255,349,424,508,536,539,602,624,661,835,881,929,957/Binding site: carbohydri  
F;492/Active site: Asp #status predicted

Query Match 28.6%; Score 1357; DB 2; Length 985;  
Best Local Similarity 34.4%; Pred. No. 1.8e-84;  
Matches 298; Conservative 142; Mismatches 290; Indels 136; Gaps 19;

QY 31 ESTVGVGVVVRSGVDNROVLTAKLDLI-KPSSVYAPDIKSLNLHVSLETSELRIRI 89  
DB 62 QSVCPGY----KASDVKHSSQGFASLELAGPCNVYDVSLSLTLYEYQAKDLNLIQI 117  
QY 90 T----DSSQQRWEI-PETVIPRAGNHSPPRSTEDGNSPENNFLADPSSDLVFTLHNT 144  
DB 118 VPTYFDASNASWYILSEELVPRP-----KASQN--ASVPQSDFFVVSWSNE 160  
QY 145 TTPGFSVSRSSGDLFDTSDDSDNTYFIKQDFQLSSALPENRSLYGIQHTKRS 204  
DB 161 PSFNFVKIRKATGVDLFTNT-----KGSTLYVEQFIEFTLLPE-EYNLYGLGERMNQ- 212  
QY 205 FRLIPGCTMTLWNADIGSENPDVNLGSHPPFYMDVRGSK-----GNE 246  
DB 213 LRLLENANLTLYADI-ADPTIDDDNIYGHAFYLDTRYKVGQNKSHYIVKSSBAEPSQE 271  
QY 247 EAGTTHGVLNLSNGMVKYEGHRTYNNVIGGVIDLIVYFAGSPENVMNQY-TBLIGRPA 305

```
Db 272 YVSYSHGVFURNHGHQBIILLRQDKLIMRTLGGSVDLTFYSGPTQAEVTKYQYSTVGLPA 331
QY 306 PMPYWFGFHQCRGYQVQVSDLEVYVDGYAKAGIPLEVMWTD:DYMDGYKDFTLDPVNP 365
Db 332 MQQNTLGFHQCRGWYNNSEFEDVLNFERPIPLEYLWADIDYMHGVRNFENDOHRFS 391
QY 366 EDKMQSFVDTLHKNQKQKYLILDPGI-----GVDSYGYTYNRMGEADVPIKR-NGEPLYL 418
Db 392 YEEGEKFLNKLHAGGRRWVPIVDGALYIPNENASDAYETDYGAKDDVFIKNPQGSLYI 451
QY 419 GSWMPGVKYPFDFLNPAAATFWSNEIKMFOEILPLDGLWIDMNELSNF----- 467
Db 452 GAVMPGYTVYEDWHRPKASDFWANELVTWNNKLHYDGWYDAAEVSSFCVSGCGTGNLSM 511
QY 468 -----TSP 470
Db 512 NPAHPPALPGERGNVYDYDEGFNITNATEASASAGAASQSAASSTTSAPYLRTTP 571
QY 471 LSSGSLDDPPYKINNSGDKRPINNKTPVATSIHFNGI SEYDAHNLVGLLEAKATHCAVV 530
Db 572 TPGVRNVDPHYVNHVQPHDLSVHAISPNSTHSDGVQVQDYVHSLYGHQGINATYHGLL 631
QY 531 DI-TGKGPFLSSTFVSSSKYTAHWTGDNAAKWEDLAYSIPIGLNFGIPGIPMGACD 589
Db 632 KYWENKRPFIARSTFGSGKAGHMGDNFSGKGMFFSISOALQFSLFGIPMGVDT 691
QY 590 GFSHDTTEELCRRWIOLGAFYFPAARDHSSLGTAQELYLWDSVASSARKVLGLRMELLP 649
Db 692 GFNGTDEELCNRMQLSAFFPYRNHNVLSATPQEPYRWASVIDATKAMNIRYAILPY 751
QY 650 LYTLMEVHVSNGNPIARLPFPSPQDTKYTIDSQFLIGKSIWVSPALQKQAVADAYFP 709
Db 752 FYTLFLAHTTGSTVMRALAWFPPDPSLAAVGTQFLVGPVNVIPVLEPVDVTVQGVFP 811
QY 710 -AGN---WEDLFNYSFVAGDSCKHVRDLTPADRVNVHVRREGSIVAMQGEALTRDARKT 765
Db 812 GVGHGEVYDYSQT-AVDKAPGVNTTISAPLGHIPFVRGSGSILPMQVEALTRDARKT 870
QY 766 PYOLLVVASLENISGELFLDDGENL 791
Db 871 PWSLLASLSNGTASQQLYLDGSEV 896

RESULT 12
T50267
probable family 31 glucosidase [imported] - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #ext_change 21-Jul-2000
C:Accession: T50267
R:Hunt, C.; Aves, S.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, December 1999
A:Reference number: Z25031
A:Accession: T50267
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Map position: 1
A:Residues: 1-995 <HUN>
A:Cross-references: EMBL:AL135522; PIDN:CAB63549.1; GSPDB:GN00066; SPDB:SPAC922.02c
A:Experimental source: strain 972h(-); cosmid c922
C:Genetics:
A:Gene: SPAC1039.11c; SPDB:SPAC922.02c
A:Map position: 1
C:Superfamily: Schwanniomyces glucan 1,4-alpha-glucosidase GAM1; sucrose/isomaltase homodimer
```

```
Db 145 WDAPLYHFQPOGDTYNFSE-----NSQPFEPWTVRSVGEVLFDTRGHK-----LIFED 195
QY 179 QFLQLSSALPENRSLNYGIGHTKRSFLIPGETMTLWADIGSEN-----DYNLYGSHPF 235
Db 196 QYIELTMMVDD-VNYYGLAE-TVHGLSLGNLRTTFP-----ANGNPTPLDRNAYGTHPF 249
QY 236 YMDVR-----GSKGNEE-AGTTHGVLLNSNGMDVKYEGHRITYNVYGVIDLIVYFAG 287
Db 250 YLEHRYTPESENLSGQSPSYTSSTHGVLMLTANGMEVLLRPNYLOYRIIGIVDLYIYVG 309
QY 288 --PSPENVMNQYTELIGRPAMPYWSFGFHQCRGYKXVSDLEVVYVQYAKAGIPLEVMW 345
Db 310 GTKQPKDTYVSVQSVGTAPAOQHWTFGFHICRMGYKNVFDLVEVKENFKNFEIPVDTFW 369
QY 346 TDIDYMDGYKDFTLDPVNVFPEDKMQSFVDTLHKNQKQKYLILDPGIV-----DSSYG 398
Db 370 SDIDYMYRDTFTVESNAFPKDKMMEFFNSLQQSNQHYVPIIDAAIYAANFINRSDVDVY 429
QY 399 TYNRGMEADVPIKR-NGEPYLGEMVPGKYPFDFLNPAAATFWSNEIKMFOEI----- 450
Db 430 PYEGVRDIDFLRNPDRSLYVGNVWPGFTTFDPFNPETTYWTECLMNLSSAAFGYNSSF 489
QY 451 -LPDLGLWIDMNELSNF----- 466
Db 490 PLPYSGLWIDMNEPTSCIGSCGDKLDQNPVHPAFILLEGEPNNVMVYMEGFEHTNASE 549
QY 467 -----ITSPSSGS----- 499
Db 550 HASAYQASVSQYATATSTVESVKATSTPLNVRPKYNINYPYALNTEQGGEDLSNLGVS 609
QY 500 ATSIHFGNISEYDAHNLVGLLEAKATHCAVVDI-TGKRPFILSRSTFVSSGKYTAHMTGD 558
Db 610 VNATYHDGTVRNLFNTYGYDQSRVTVDSLTSEPNVRPFILSRSTFVSGSKYAAHMLGD 669
QY 559 NAAKWEEDLAYSIPIGLNFGIPGIPMGVADICGPHSDTTEELCRWIOGAYPPARDHSS 618
Db 670 NYSLSNNMIFIPGALTENMVGLEPMVGVADGCPMGNTDBELCSRMMALGAFLPFYRNHS 729
QY 619 LGTAQELYLWDSVASSARKVLGLRMELLPHTLYTLMEVHVSNGNPIARLPFPSPQDTKT 678
Db 730 LOSISQEPYRWESVAESRCAMNIRYSLPYWYTLMEASSQGLPLRPLRPFEPNEPSL 789
QY 679 YEIDSQFLIGKSIWVSPALQKQAVADAYFPAGN---WFDLFNYSFVAGDSCKHVRDLT 735
Db 750 ANADRQFMVGSALLVTPLEPNVDYVRGVFGDNSTIWDYDHK-VIYRQHNENITLSA 848
QY 736 PADHVNVRREGSIVAMQGEALTRDARKTYPQLLYVASRLENISGELFLDDGGENLMA 795
Db 849 PLTHINVAIRGNIIIPMQKPSLTTHETKQNPYDLVALDSDRKACGSLYVDGVSIOQ--- 905
QY 796 GGNRDMTL-VKPRCYVTGKSVLRS-----EWNVPEVASKMKWIGKVTFGV----- 842
Db 906 ---QESTLFVKP-VANGDSLSIESYCDLQVHEP-----LSKTIIGLPCAPIGV 950
QY 843 -FENVE 847
Db 951 YFEGVQ 956

RESULT 13
JN0102
Glucan 1,4-alpha-glucosidase (EC 3.2.1.3) GAM1 precursor - yeast (Schwanniomyces occiden-
N:Alternate names: acid maltase; Glucoamylase
C:Species: Schwanniomyces occidentalis
C:Date: 10-Mar-1994 #sequence_revision 05-Apr-1995 #text_change 18-Jun-1999
C:Accession: JN0102
R:Dohmen, R.J.; Strasser, A.W.M.; Dahlems, U.M.; Hollenberg, C.P.
Gene 95, 111-121, 1990
A:Title: Cloning of the Schwanniomyces occidentalis glucoamylase gene (GAM1) and its c-
A:Reference number: JN0102; MUID:91071592; PMID:1979298
A:Accession: JN0102
A:Molecule type: mRNA
A:Residues: 1-958 <DOH>
```

```

Db      843 AESRQNPFGGLIVALDNDGKAQGSGLYLDGSLVYDSS-----LLVSVF-----SVSDN 889

QY      820 SEVVNPYASAKMWSIGKVTFG-----FENVENVKTYEVRT 856

Db      890 TLSASPGDYKADQPIANVTILGVGHKPKSVKFNANVDFYTKST 935

RESULT 14
T38598
probable family 31 glycosyl hydrolase (alpha glucosidase) precursor - fission yeast
C:Species: Schizosaccharomyces pombe
C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
C:Accession: T38598; S62559; T38911
R:Pearson, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, November 1995
A:Reference number: Z21801
A:Accession: T38598
A:Molecule type: DNA
A:Residues: 1-385 <PEA>
A:Cross-references: EMBL:Z67961; PIDN:CAA91887.1; GSPDB:GN00066; SPDB:SPAC30D11.01c
A:Experimental source: strain 972h-; cosmid c30D11
R:Pearson, D.; Churcher, C.M.
submitted to the EMBL Data Library, November 1995
A:Reference number: S62559
A:Accession: S62559
A:Molecule type: DNA
A:Residues: 1-384 <PE2>
A:Cross-references: EMBL:Z67961
R:Pearson, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z21817
A:Accession: T38911
A:Molecule type: DNA
A:Residues: 352-993 <PR3>
A:Cross-references: EMBL:Z69728; PIDN:CAA93572.1; GSPDB:GN00066; SPDB:SPAC56F8.01
A:Experimental source: strain 972h-; cosmid c56F8
C:Genetics:
A:Gene: SPDB:SPAC30D11.01c; SPDB:SPAC56F8.01
A:Map position: 1L
C:Superfamily: Schwanniomycetes glucan 1,4-alpha-glucosidase GAM1; sucrase/isomaltase
F:1-65-384/Domain: sucrase/isomaltase homology (fragment) <SIM>

Query Match      28.3%; Score 1345.5; DB 2; Length 993;
Best Local Similarity 33.6%; Pred. No. 1.le-83;
Matches 308; Conservative 151; Mismatches 302; Indels 157; Gaps 24;

QY      38 GYVVRSGVDSNRQVLTKLDLIKPS--VYAPDIKSLNHVSLETSERLRIRITDSSQOR 96
Dc      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84 GYQARNISEYS--YGVLAILELAGDACYAGTDYPLLNVSDTERVHISISDLNQIQ 141
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY      97 WEIPETVIPAGNHSPPRSTEEGGNSP-----ENNFLADPPSDLVFTLHNTTPFGFSVS 152
Dc      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
142 FQL-----SNRRDVMADPLFYRRSSNFGNLQYNFSF---NTDPDFEFWIT 182
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY      153 RRSSGGDILPDTSPDSSDNTYRIFKDFQLQSLALPENRSNLNGICEHTKRSFRLIPGET 212
Dc      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
183 RIADQVLFDTGRNP-----LIFEQYIELTNMYED--YNYVGL--SGSQSQSFRLLGNLIT 234
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY      213 MTLWNADIGSENPVNLYGSHPPYMDVR---GSKGNEEAGTTHGYLLNSNGMDYKYG 268
Dc      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
235 KTFW-ATGYSDSPAANNYGSHPPYMEQRVIPGTT-NTVTSASHGVLMSSNGMEVLLRS 292
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY      269 HRTYNYVIGVIDLYYFAGP--SPENVNMQYTELICRPAPMPYWSFGHCQRCVYKQVSD 326
Dc      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
293 TYIKRMIGIGIDLFYVSGTSTSPKTYIQOYVOSIGTPTMQPYWSLGFQMSRWGYKTLSD 352
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY      327 LEYVVDGYAKGIFLEVMWMTDIDYMDGYKDFLDPVNFPEDKKQSFVDTLHKNGQKYVLI 386
Dc      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
353 LINRSLVNASNPTSGFWNDIDYNSFEFTFYNSAFPNQTLDFRSLDSESHQYVPV 412
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY      387 LDPGI-----GVDSSYGTGYNRGMEADVFIKR--NCEPYLGEVWPQKGYFPDFLNPAAAT 438
Dc      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
413 LDPAIYAANPNKSDRTYYPYSGFDNFIPIKNPGSAYGVNMAWPGVYVYPDFTNPAVLQ 472
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:20:43 ; Search time 108.36 seconds  
(without alignments)  
1393.955 Million cell updates/sec

Title: US-10-043-418-4

Perfect score: 4751  
Sequence: 1 MSSLHWPNIPIVVVFFSL.....EVSLSLVGKKFEMRLRLT 902

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 629382 seqs, 167460630 residues

Total number of hits satisfying chosen parameters: 629382

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10C\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4751	100.0	902	14	US-10-043-418-4
2	2830.5	59.6	903	14	US-10-043-418-3
3	2719.5	57.2	914	14	US-10-043-418-2
4	2451	51.6	877	14	US-10-043-418-1
5	1954.5	41.1	707	12	US-10-032-189-125
6	1360	28.6	1827	9	US-09-819-247-2
7	877.5	18.5	912	12	US-10-032-189-44
8	876.5	18.4	914	12	US-10-032-189-40
9	876	18.4	966	12	US-10-032-189-120
10	875.5	18.4	914	12	US-10-032-189-42
11	862	18.1	943	12	US-10-032-189-123
12	862	18.1	944	12	US-10-032-189-122
13	851.5	17.9	966	12	US-10-032-189-121
14	837.5	17.6	967	12	US-10-032-189-38
15	832	17.5	944	12	US-10-032-189-124

16	761	16.0	565	15	US-10-102-806-557	Sequence 557, Appli
17	713.5	15.0	693	12	US-10-228-063-5	Sequence 5, Appli
18	713.5	15.0	712	12	US-10-228-063-27	Sequence 27, Appli
19	713.5	15.0	718	12	US-10-228-063-26	Sequence 26, Appli
20	713.5	15.0	718	12	US-10-228-063-36	Sequence 36, Appli
21	653.5	13.8	788	15	US-10-156-761-14497	Sequence 14497, A
22	601.5	12.7	235	9	US-09-734-569-132	Sequence 132, App
23	494.5	10.4	1070	10	US-09-280-197-6	Sequence 6, Appli
24	494.5	10.4	1070	10	US-09-280-197-4	Sequence 4, Appli
25	479.5	10.1	1066	10	US-09-280-197-5	Sequence 5, Appli
26	479.5	10.1	1066	10	US-09-280-197-3	Sequence 3, Appli
27	420.5	8.9	642	15	US-10-156-761-10107	Sequence 10107, A
28	404	8.5	188	9	US-09-734-569-134	Sequence 134, App
29	365	7.7	1091	10	US-09-280-197-2	Sequence 2, Appli
30	365	7.7	1091	10	US-09-280-197-1	Sequence 1, Appli
31	358.5	7.5	1088	10	US-09-423-126-2	Sequence 2, Appli
32	358.5	7.5	1088	10	US-09-423-126-1	Sequence 1, Appli
33	356.5	7.5	1092	10	US-09-423-126-5	Sequence 5, Appli
34	334	7.0	199	9	US-09-734-569-130	Sequence 130, App
35	333.5	7.0	570	10	US-09-423-126-6	Sequence 6, Appli
36	146	3.1	1129	15	US-10-156-761-8946	Sequence 8946, Ap
37	133.5	2.8	1024	15	US-10-213-990-30	Sequence 30, Appli
38	131.5	2.8	317	10	US-09-280-197-9	Sequence 9, Appli
39	123.5	2.6	846	9	US-09-815-242-13904	Sequence 13904, A
40	123	2.6	1151	11	US-09-984-130-103	Sequence 103, App
41	123	2.6	1151	12	US-09-836-353A-103	Sequence 103, App
42	122	2.6	1038	15	US-10-156-761-13515	Sequence 13515, A
43	121.5	2.6	5636	12	US-10-032-189-128	Sequence 128, App
44	119.5	2.5	1900	10	US-09-801-368-390	Sequence 390, App
45	119	2.5	1300	9	US-09-815-242-4903	Sequence 4903, Ap

ALIGNMENTS

RESULT 1  
US-10-043-418-4  
; Sequence 4, Application US/10043418  
; Publication No. US20020184662A1  
; GENERAL INFORMATION:  
; APPLICANT: Henson, Cynthia A.  
; APPLICANT: Muslin, Elizabeth H.  
; APPLICANT: Clark, Suzanne E.  
; TITLE OF INVENTION: Modified barley alpha-glucosidase  
; FILE REFERENCE: 960296.97486  
; CURRENT APPLICATION NUMBER: US/10/043.418  
; CURRENT FILING DATE: 2002-06-25  
; PRIOR APPLICATION NUMBER: 60/260,787  
; PRIOR FILING DATE: 2001-01-10  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 902  
; TYPE: PRT  
; ORGANISM: Arabidopsis  
US-10-043-418-4

Query Match		100.0%;	Score 4751;	DB 14;	Length 902;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 902;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MSSLHWPNIPIVVVFFSLRSSQVLEEEESTVVGYYVRSVGVDSNRQVLTAKLDLI	60		
Db	1	MSSLHWPNIPIVVVFFSLRSSQVLEEEESTVVGYYVRSVGVDSNRQVLTAKLDLI	60		
Qy	61	KPSVYAPDIKSLNLHVSLETSELRIRITDSSQORWEIPETVIPRAGNHSRPFSTEED	120		
Db	61	KPSVYAPDIKSLNLHVSLETSELRIRITDSSQORWEIPETVIPRAGNHSRPFSTEED	120		
Qy	121	CGNSPENNLADPSSDLVFTLHNTTPEGFSVRSRSGDILFDTPSDSDSNTYFIKQDF	180		
Db	121	CGNSPENNLADPSSDLVFTLHNTTPEGFSVRSRSGDILFDTPSDSDSNTYFIKQDF	180		



```
QY 181 LQLSSALPENRSLNYGIGETKRSFRLLPGETMTLWNADIGSENPDVNLVYSGSHPFYMDVR 240
D 181 LQLSSALPENRSLNYGIGETKRSFRLLPGETMTLWNADIGSENPDVNLVYSGSHPFYMDVR 240
QY 241 GSKGNEEAGTTHGVLLNSNGMDVYKYECHRTTYNVIGVVDLYVYFAGSPSEMVNMYTEL 300
D 241 GSKGNEEAGTTHGVLLNSNGMDVYKYECHRTTYNVIGVVDLYVYFAGSPSEMVNMYTEL 300
QY 301 IGRPAPMYWFGFHQCRGYKYNVSDLEYVVDGYAKAGIPLVNMWTDIDYMDGYKFTLD 360
D 301 IGRPAPMYWFGFHQCRGYKYNVSDLEYVVDGYAKAGIPLVNMWTDIDYMDGYKFTLD 360
QY 361 PVNFPEDKMQSFVDTLHKNGQKYYVLLDPGIGVDSSYGTNRGMEADVFKRNGEYVIGE 420
D 361 PVNFPEDKMQSFVDTLHKNGQKYYVLLDPGIGVDSSYGTNRGMEADVFKRNGEYVIGE 420
QY 421 VMPGKVYFPDFLNPAAATFWSNEIKMFQELPLDGLWIDMNELSNFTITSLSSGSSLDOP 480
D 421 VMPGKVYFPDFLNPAAATFWSNEIKMFQELPLDGLWIDMNELSNFTITSLSSGSSLDOP 480
QY 481 PYKINNSGDKRPPINNKTPVATSIHFGNISEYDAHNLYGELLEAKATHQAVVDITGKRPFIL 540
D 481 PYKINNSGDKRPPINNKTPVATSIHFGNISEYDAHNLYGELLEAKATHQAVVDITGKRPFIL 540
QY 541 SRSTFVSSGKYTAHMTGDNAAKWBEDLAYSIPGILNFGFLGIPMVGADICGFSHDTTEELC 600
D 541 SRSTFVSSGKYTAHMTGDNAAKWBEDLAYSIPGILNFGFLGIPMVGADICGFSHDTTEELC 600
QY 601 RRTWQLGAFYPPARDHSLGRTARQELYLWDSVASSARKVGLRMRLPHLYTYLMYEAHVS 660
D 601 RRTWQLGAFYPPARDHSLGRTARQELYLWDSVASSARKVGLRMRLPHLYTYLMYEAHVS 660
QY 661 GNPIARPLFFSFPDQTKTYEIDSOFLIGKSIWVSPALKOGAVAVDAVFPAGNWFDLNYS 720
D 661 GNPIARPLFFSFPDQTKTYEIDSOFLIGKSIWVSPALKOGAVAVDAVFPAGNWFDLNYS 720
QY 721 FAVGDSGKHVRLDTPADHVNHVHREGSIVAMQGEALTTRDARKTPYQLLVASRLNLS 780
D 721 FAVGDSGKHVRLDTPADHVNHVHREGSIVAMQGEALTTRDARKTPYQLLVASRLNLS 780
QY 781 GELFLDGENLRMGAGGGRDWTLLVKRCYVTKGSVVLSEVNVPEYASOKKWSIGKVIF 840
D 781 GELFLDGENLRMGAGGGRDWTLLVKRCYVTKGSVVLSEVNVPEYASOKKWSIGKVIF 840
QY 841 VGFENVENVKTYYEVRTSERLSPRISLIKTYSNDNDPRFLSVEVSKLSLLVGKKFEMRLR 900
D 841 VGFENVENVKTYYEVRTSERLSPRISLIKTYSNDNDPRFLSVEVSKLSLLVGKKFEMRLR 900
QY 901 LT 902
D 901 LT 902
RESULT 2
US-10-043-418-3
; Sequence 3, Application US/10043418
; Publication No. US20020184662A1
; GENERAL INFORMATION:
; APPLICANT: Henson, Cynthia A.
; APPLICANT: Muslim, Elizabeth H.
; APPLICANT: Clark, Suzanne E.
; TITLE OF INVENTION: Modified barley alpha-glucosidase
; FILE REFERENCE: 960296.97486
; CURRENT APPLICATION NUMBER: US/10/043,418
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: 60/260,787
; PRIOR FILING DATE: 2001-01-10
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3
; LENGTH: 903
; TYPE: PRT
; ORGANISM: Spinach
```

US-10-043-418-3

```
Query Match          59.6%; Score 2830.5; DB 14; Length 903;
Best Local Similarity 59.6%; Pred. No. 8.7e-255;
Matches 540; Conservative 138; Mismatches 205; Indels 23; Gaps 8;

QY 8 PNIFIVVVVFSLSR---SSQVVLREEESTVVGYYVVRSGVDS-NROVLTAKLDLIKPS 63
D 6 PSLAIGIILLVFLQYLVAGISTENSEDPEGVYGYGYKVKSVKDSGTRSLTALPQLVKN 65
QY 64 SVYAPDIKSLNHLVSETSERLRIRITDSSQORWEIPETVIPRAGNHSP-----RRFST 117
D 66 SVYGPDIQLLSITASLESNDRLRVITDAKRRWEIPDNILHR---HQPPPPPHSLSL 122
QY 118 EDGGSNPNENN----FLADPSSDLVFTLHNTTPFGFVSRRSSGDIILFDTSPDSDSN 173
D 123 VRTLLSSPTTNRKILLSHPNSDLTFSLINTTPFGFTISRKSTHDLVDFATPDPTNPT 182
QY 174 FIFKDOFLQLSALPENRSLNYGIGETKRSFRLLPGETMTLWNADIGSENPDVNLVY 233
D 183 LIFIDQYUHLTSSUPGTTRAHIYGLGHSKPTFQLAHNQTLTMRADIPSSNPVNLVY 242
QY 234 PFYMDVRSGKNEEAGTTHGVLLNSNGMDVYKYECHRTTYNVIGVVDLYVYFAGSPSE 293
D 243 PFYMDVRSS---PVAGSTHGVLNLSNGMDVEYTGNRITYKVIIGIIDLFFAGSPSQ 299
QY 294 MNQYTELIGRPAPMYWFGFHQCRGYKYNVSDLEYVVDGYAKAGIPLVNMWTDIDY 353
D 300 VEQFTRVTRIGRPAPMYWFGFHQCRGYKYNVSDLEYVVDGYAKAGIPLVNMWTDID 359
QY 354 YKDFTLDPVNPEDKMQSFVDTLHKNGQKYYVLLDPGIGVDSSYGTNRGMEADVFKR 413
D 360 YKDFTLDPVNPEDKMQSFVDTLHKNGQKYYVLLDPGIGVDSSYGTNRGMEADVFKR 419
QY 414 GEPYLGEMVPGKVYFPDFLNPAAATFWSNEIKMFQELPLDGLWIDMNELSNFTIT 473
D 420 GKPYLGSMVPGKVYFPDFLNPAAATFWSNEIKMFQELPLDGLWIDMNELSNFTIT 479
QY 474 GSSLDQDPYKINNSGDKRPPINNKTPVATSIHFGNISEYDAHNLYGELLEAKATH 533
D 480 GSTLDNPPYKINNSGDKRPPINNKTPVATSIHFGNISEYDAHNLYGELLEAKATH 539
QY 534 GKRPFILSRSTFVSSGKYTAHMTGDNAAKWBEDLAYSIPGILNFGFLGIPMVGADI 593
D 540 EKRPVLSRSTFVSSGKYTAHMTGDNAAKWBEDLAYSIPGILNFGFLGIPMVGADI 599
QY 594 DTTEELCRRTWQLGAFYPPARDHSLGRTARQELYLWDSVASSARKVGLRMRLPH 653
D 600 NTTEELCRRTWQLGAFYPPARDHSLGRTARQELYLWDSVASSARKVGLRMRLPH 659
QY 654 MYEAHVSGNPIARPLFFSFPDQTKTYEIDSOFLIGKSIWVSPALKOGAVAVDAV 713
D 660 MYEAHVSGNPIARPLFFSFPDQTKTYEIDSOFLIGKSIWVSPALKOGAVAVDAV 719
QY 714 FDLFNYSFVAGDSGKHVRLDTPADHVNHVHREGSIVAMQGEALTTRDARKTPYQL 773
D 720 FDLFDYTRSVTASTGRVYTLTAPDHVNHVHREGSIVAMQGEALTTRDARKTPYQL 779
QY 774 SRLENISGELFLDGENLRMGAGGGRDWTLLVKRCYVTKGSVVLSEVNVPEYAS 833
D 780 SDGASFGELFLDGENLRMGAGGGRDWTLLVKRCYVTKGSVVLSEVNVPEYAS 837
QY 834 SIGKVTFVGFENVENVKTYYEVRTSERLSPRISLIKTYSNDNDPRFLSVEVSK 893
D 838 VIDKVTILGLRKGTINGYVTRTCAVTRKDGKSKLSTPDRKG-EFIVAEISGLN 896
QY 894 KFEMRL 899
D 897 EFKLVL 902
```

RESULT 3  
US-10-043-418-2

```
; Sequence 2, Application US/10043418
; Publication No. US20020184662A1
; GENERAL INFORMATION:
; APPLICANT: Henson, Cynthia A.
; APPLICANT: Muslin, Elizabeth H.
; APPLICANT: Clark, Suzanne E.
; TITLE OF INVENTION: Modified barley alpha-glucosidase
; FILE REFERENCE: 960296.97486
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: 60/260,787
; PRIOR FILING DATE: 2001-01-10
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 914
; TYPE: PRP
; ORGANISM: Sugar beet
; US-10-043-418-2

Query Match      57.2%; Score 2719.5; DB 14; Length 914;
Best Local Similarity 57.0%; Pred. No. 2.1e-244;
Matches 516; Conservative 146; Mismatches 219; Indels 25; Gaps 7;

QY 8 PNIFIV--VVFFSRSSQVLEEEESTVVGYYVRSVGVVD-SNRQVLTAKLDLIKPS 64
DB 12 P'LAVVLPVLCMVVGGATTSKNDNGEAGYGYQVKNKAVDNGSTGSLTALLQJRN 71
QY 65 VYAPDIKSLNHLVSLTSELRIRITDSSQORWEIPETVIR-----AGNHSR 113
DB 72 VYGPDIHFLSTASFEEDDTLRIRFTDANNRWEIPNEVLPRPPPPPPPPSLSSQHL 131
QY 114 RSTEEEDGSGENNPLADPSSDLVFTLHNTTFFGFSVRSRSGDILPOTSPDSSDNTY 173
DB 132 PIPQ-----NQTTTVLSHPHSDLAFTLHTTFFGFTYRKSTHVDLPATPISNPT 186
QY 174 FIFKQOFLSALLPENRNLVIGIEHTKRSRLIPGETMTLWNADIGSENPVNLYGSH 233
DB 187 LIYKQYQLSSLSQAQAHLYGLGHEHTKPTQLAHNQILTLWNADIASFNRDLNLYG 246
QY 234 PYMDVRSGKNEEAGTTHGVLLNSGMDVKYEGHRITVNYVIGVLDLYVAGSPPEW 293
DB 247 PYMDVRSS---PMVGSTHGVLNSGMDVETGDRITVYKIGIILDIYPAGTPEW 303
QY 294 MNQYTELIGRPAMPYWSFGFHQCRYGKNSDLEYVVDGYAKGIPLEVMTDIDYMDG 353
DB 304 LQOYTKLIGRPAMPYWSFGFHQCRGWEDVNEITVVDKYAEARIPLEVMTDIDYMD 363
QY 354 YKDFLDPVNFPEDKQSFVDTLHNGQKQVLIPLPGIGVDSYSGTYNRGMEADVFIRN 413
DB 364 FKDFLDPVHFPLDKMQQFVTKLHRNGQRYVPILOPGINTNKSQYTFIRMGQSNV 423
QY 414 GBPYLEGVWFGKYVFPDFLNPAAATFWSNEIKMFOELPLDGLGWLDMNLSNFI 473
DB 424 GNPYLGSVWFGPYVFPDFLNPAAATFWSNEIKRFRDILPIDGIMDMNLSNFI 483
QY 474 GSSLDPPYKINNSGDKRPINNKTVPATSIHFGNTSEYDAHNLGLLEAKATHQAV 533
DB 484 GSTLDNPPYKINNSGGRVINSKTIPTATAMHGNVTEYNAHNLGLLESQATREAL 543
QY 534 GKRPFTLSSTFVSSGKYTAHMTGNAKWEIDLAYSIPGLNFGIPGTPMVCADICG 593
DB 544 TRGPFLLSSTFAGSGKYTAHMTGDAARWDLQVSIPTMLNFGIPGTPMVCADIC 603
QY 594 DTTEELCRWIOLGAFYPTFARHSSLGTAQELYLWDSVASSARKVGLRMLLPHLY 653
DB 604 SITEELCCRWIOLGAFYPTSRHSARDTHQELYLWESVAASARTVGLRYELLPY 663
QY 654 MYEAWGNPIARLPFFSQDTKTYEIDSQFLGKSIWSPALQKQAVDVPFAGNW 713
DB 664 MYDANLRGSIARPLSFTFFDDVATYGISSQFLIGRGIWSPVLPQFGSSVNAK 723
QY 714 FOLFNYSFAVGSDSGKHVRLDTPADHVNHVHREGSIVAMQGEALTRDARKTP 773
```

```
DB 724 VLSNYSVSVSAGYVVSLSAPPDHNHIIHGNIVAMQGEAMTTQARSTFPHLLV 783
QY 774 SRLNISGELFDGGENLRMGAGGGRNDWTLVFRCYVTGKSVLRSSEVNVNYSKMK 833
DB 784 SDHVASTGELFDGEMIDTGGPGK--WTLVRFPAESGINNLTISSEVNVNYSK 841
QY 834 SICKVTFVGFENVKTEVTRSERLSPRISLIKTVDNDPRLSVSEVSKLSLVCK 893
DB 842 VMDKIITLGLKRRVKIKEYTVQDAGAIKVGGLGRRTSSHNQGGFFVSV-IS 900
QY 894 KFEKRL 899
DB 901 AFKLEL 906

RESULT 4
US-10-043-418-1
; Sequence 1, Application US/10043418
; Publication No. US20020184662A1
; GENERAL INFORMATION:
; APPLICANT: Henson, Cynthia A.
; APPLICANT: Muslin, Elizabeth H.
; APPLICANT: Clark, Suzanne E.
; TITLE OF INVENTION: Modified barley alpha-glucosidase
; FILE REFERENCE: 960296.97486
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: 60/260,787
; PRIOR FILING DATE: 2001-01-10
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 877
; TYPE: PRP
; ORGANISM: Barley
; US-10-043-418-1

Query Match      51.6%; Score 2451; DB 14; Length 877;
Best Local Similarity 55.2%; Pred. No. 2.4e-219;
Matches 471; Conservative 136; Mismatches 191; Indels 56; Gaps 13;

QY 69 DIKSLNHLVSLTSELRIRITDSSQORWEIPETVIRAG-----NHSPPRFSTE 123
DB 59 DVORLAVYASLETDSRLVRITDADHPRWEVPODIIPRAPGVDVLHDAPPASAP 116
QY 124 SPENFLADPSSDLVFTLHNTTFFGFSVRSRSGDILPOTSPDSSDNTYFIFK 183
DB 117 ----RVLSPAGSDLVLTVH--ASPFRTVSRSTGDTLFDTPAG-----LV 164
QY 184 SSALPENRNLVIGIEHTKRSRLIPGETMTLWNADIGSENPVNLYGSHPFYMD 243
DB 165 TSALPAGRSYLGIEHTKSSFLRHNDSPFLWNADIGASYVDVNLVYSGHPFYM 222
QY 244 GNEAGTTHGVLLNSGMDVKYEGHRITVNYVIGVLDLYVAGSPPEWVMOYTEL 303
DB 223 ---PGTAHGVLLNSGMDVLYGSSYVTVYKVGVLDFYFFAGPNPLAVVDOYT 278
QY 304 PAPMPYWSFGFHQCRYGKNSDLEYVVDGYAKGIPLEVMTDIDYMDGYKFTD 363
DB 279 PAPMPYWSFGFHQCRYGKNSDLEVRVARYAKARIPLEVMTDIDYMDGFKD 338
QY 364 FPKQMSQFVDTLHNGQKQVLIPLDIPGIGV---DSSYGTYNRGMEADVFIK 420
DB 339 FTAELRPFVDRLHRNAQKVLILDPGIRVDPIDATYGTGVRGQQDIPLKRN 398
QY 421 VWGKYVFPDFLNPAAATFWSNEIKMFOELPLDGLGWLDMNLSNFI-SPLSSG 479
DB 399 VWGKYVFPDFMHPAAAEFWAREISLFRITPVYDGLWDMNEISNFYNPEPM 455
QY 480 PPKYKINNSGDKRPINNKTVPATSIHFGNTSEYDAHNLGLLEAKATHQAV 539
DB 456 PPRINNDGTGRPINNKTVPLAVHYGGVTEYBEHNLFGLLLEARATGRVLR 515
```

```

Qy 540 LSRSTFVSSGKYTAHTWGDNAKAWEDLAYSPGLNPLGFGIPMGVADICGFSHDTTEL 599
Db 516 LSRSTFVSGRYTAYWTGDNAAWTGDLRYSTINTWLSGLFGFPMICADICGFNGTTEL 575
Qy 600 CRRWQIQAQFPPFARDHSSILGTARQELYLWDSVASSARKVLGRMLLPHLYTLMAEHV 559
Db 576 CRRWQIQAQFPPFARDHSSILGTARQELYLWDSVASSARKVLGRMLLPHLYTLMAEHV 559
Qy 660 SGNIARPLFFSPQDTKYEIDSOFLIGKSIWSPALKQGAVDAYFPAGNWFELNY 719
Db 636 TGAPIARPLFFSPYDVAQVDRQFLGRGLVSPVLEPGPTTVDAYFPAGRWYLYDY 695
Qy 720 SFVAGDSGKHVLDTPADHVNHVHREGSIVANQGEALITRDARKTPYQLLVASLENI 779
Db 696 SLAVATETGKHVLPAPADTVNHLTGITLPLQSSALITTSRAPRTAFHLLVALAEDGTA 755
Qy 780 SGELFLDDGELNLRGAGGGRNDWTLVKFRVCYVTKG--SVVLSEVVNPEYASQKWSICK 837
Db 756 SGYLFDDGDSPEYGR---RSDMSWRFNYKIPNNKGAIKVSEVHNSYAQSRTLVISK 812
Qy 838 VTFVGFENENVK--TYEVTSRSLRSPRISLIKTVDNDPRLS-----VEVSKLS 888
Db 813 VVLGHRSPAAPKLITVHNSAE-----VEASSAGTRYQAGGLGVVAH:GGLS 862
Qy 889 LLVGKXKPEMRRLT 902
Db 863 LVUGEFEKLVAMS 876

RESULT 5
US-10-032-189-125
; Sequence 125, Application US/10032189
; Publication No. US20030170630A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook II, John P
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Zerbussen, Bryan D
; APPLICANT: Patturajan, Meera
; APPLICANT: Grosse, William M
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Shimkets, Richard A
; APPLICANT: Grosse, William M
; APPLICANT: Szekeres, Edward S
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Li, Li
; APPLICANT: Casman, Stacie J
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Gorman, Linda
; APPLICANT: Gangolli, Esha A
; APPLICANT: Fernandes, Elma R
; APPLICANT: Rieger, Daniel K
; APPLICANT: Edinger, Shomit R
; APPLICANT: Gunther, Erik
; APPLICANT: Millet, Isabelle
; APPLICANT: Sciore, Paul
; APPLICANT: Ellerman, Karen
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glennda
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-228
; CURRENT APPLICATION NUMBER: US/10/032,189
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,495
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/258,171
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/269,940
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 60/274,192

```

```

; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/277,826
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/279,840
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/282,981
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/283,656
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/309,247
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/311,754
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/313,331
; PRIOR FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 260
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 125
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Glycosyl
; OTHER INFORMATION: hydrolases family 31
US-10-032-189-125

Query Match 41.1%; Score 1954.5; DB 12; Length 707;
Best Local Similarity 54.0%; Pred. No. 3.9e-173;
Matches 388; Conservative 100; Mismatches 193; Indels 37; Gaps 16;

```

```

Qy 135 SDLVFTLHNTTTPFGFSVSRSSGDIPLDTSFSDSDSNTY--FIFKQDFLQLSALPENRS 192
Db 13 SDLYDVLSNCGPGFEVIRKSTGDLVFDT-----TFGPLVFSQFLQLSLFS--E 62
Qy 193 NLYGIGHTKSRPLIFGETMTLWADIGSENPDVNLVYGHSHPPYMDVRGSKGNEEAGTTH 252
Db 63 YIYGLGEHAHKLPRDNTNETYTLNWRDVGYPGSDNNLYGSHPPYMSL-----EDSGNAH 116
Qy 253 GVLLNSNGMDVKY-EGHRTYNNVIGVIDLYVYFAGSPSEMVMNQYTELIGRAPMPYWS 311
Db 117 GVFLLSNAMEVDIGPGPALTYYRIGGILDFYFLGPTPEDVLQOQYTELIGRALPPYWS 176
Qy 312 FGFHQCRYGKNVSDLEYVDGYAKAGIPLEVMTDIDYMDGYKDFTLDPVNPFPDMQMS 371
Db 177 LGFHLCRWGYTNVSEVKTVDGMRKANIPLDVQWLDIDYMDGYKDFTFDVPVFPGP--ED 234
Qy 372 FVDTLHKNGOKYVLIDPGIGVDG--SYGTYNRGMEADYFIKR--NGBPYLGCEVMPGVYFP 429
Db 235 FVKLHAKGQKYVVILOPAISVDSASYYPYERKEGKGVFNPNPNSGSIYIGCEVMPGYTAPP 294
Qy 430 DFLNPAATFMSNEIKMFQELPLDGLWIDMNELSNFITSFLSSGSLDDPPYKINNSGD 489
Db 295 DFTNPEARKWADLIDKFHDSLPFDGIWIDMNEPSSP--SEPGPNDNLNYPY-APNDGD 352
Qy 490 KRPNNKTVPATSIHFQNISEYDAHNLVGLLEKATHQAVDIT--GKRPFILSRSTFVSS 548
Db 353 -GPLSSKTCMDAVHYGVEHYDVHNLVGLSEAKATVEALKVKVTGGKRPVLSRSTFAGS 411
Qy 549 CKYTAHWTGDNNAKWEIDLAYSIPGLNPLGFGIPMGVADICGFSHDTTELCHRRWOLGA 608
Db 412 GRVAGHTWGNTASWDDLYSIPGVLSFNLGFPFVGADICGFNGNTTEELCYRMWOLGA 471
Qy 609 FYFPAHSSIGTARQELYLWDSVASSA-RKVGLGRMLLPHLYTLMAEYAHVSGNPIARP 667
Db 472 FYFSPRNNHLLGTIPQEPWLFDSVAAEASRKALNRLVTLPLYLTLFHEAHVSGLPWMP 531
Qy 668 LFFSPQDTKYEIDSOFLIGKSIWSPALKQGAVDAYFPAGNWFELNYFSAVAGDS 727
Db 532 LFFFPDDDAETDYIDROFLMGSAALLVAPVLEPGATSVKAYLPGRWYDL--YTCAGEASR 589
Qy 728 GKHYRLTDPADHVNHVHREGSIVANQGEALITRDARKTPYQLLVASLENI:SGELPLDD 787
Db 590 GGNVTLAPLDKIPVHVVRGGSIIPTQEPALTTTESRDNPFHLLVALDDNGTASGEUYLDD 649

```



```

; PRIOR APPLICATION NUMBER: 60/313,331
; PRIOR FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 260
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 44
; LENGTH: 912
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-032,189-44

```

Query Match 18.5%; Score 877.5; DB 12; Length 912;  
Best Local Similarity 27.7%; Pred. No. 2.5e-72;  
Matches 253; Conservative 152; Mismatches 326; Indels 183;

[illegible]

```

Qy      896 KLSLLVGKKFEMRL 899
      |||| : :||:
Db      898 KLSLNATDMEVRI 911

RESULT 8
US-10-032-189-40
; Sequence 40, Application US/10032189
; Publication No. US20030170630A1

```

## RESULT a

```

US-10-032-189-40
: Sequence 40, Application US/10032189
: Publication No. US20030170630A1
: GENERAL INFORMATION:
: APPLICANT: Alsobrook II, John P
: APPLICANT: Tchernev, Velizar T
: APPLICANT: Liu, Xiaohong
: APPLICANT: Spytek, Kimberly A
: APPLICANT: Zernhusen, Bryan D
: APPLICANT: Patturajan, Meera
: APPLICANT: Grosse, William M
: APPLICANT: Lepley, Denise M
: APPLICANT: Burgess, Catherine E
: APPLICANT: Shinkets, Richard A
: APPLICANT: Grosse, William M
: APPLICANT: Szekeres, Edward S
: APPLICANT: Vernet, Corine A.M.
: APPLICANT: Li, Li
: APPLICANT: Casman, Stacie J
: APPLICANT: Boldog, Ferenc L
: APPLICANT: Gorman, Linda
: APPLICANT: Gangolli, Esha A
: APPLICANT: Fernandes, Elma R
: APPLICANT: Rieger, Daniel K
: APPLICANT: Edinger, Shlomit R
: APPLICANT: Gunther, Erik
: APPLICANT: Millet, Isabelle
: APPLICANT: Sciore, Paul
: APPLICANT: Ellerman, Karen
: APPLICANT: MacDougall, John R
: APPLICANT: Smithson, Glenda
: FILE OF INVENTION: Proteins and Nucleic Acids Encoding Same
: FILE REFERENCE: 21402-228
: CURRENT APPLICATION NUMBER: US/10/032,189
: CURRENT FILING DATE: 2001-12-21
: PRIOR APPLICATION NUMBER: 60/257,495
: PRIOR FILING DATE: 2000-12-21
: PRIOR APPLICATION NUMBER: 60/258,171
: PRIOR FILING DATE: 2000-12-20
: PRIOR APPLICATION NUMBER: 60/269,940
: PRIOR FILING DATE: 2001-02-20
: PRIOR APPLICATION NUMBER: 60/274,192
: PRIOR FILING DATE: 2001-03-08
: PRIOR APPLICATION NUMBER: 60/277,826
: PRIOR FILING DATE: 2001-03-22
: PRIOR APPLICATION NUMBER: 60/279,840
: PRIOR FILING DATE: 2001-03-29
: PRIOR APPLICATION NUMBER: 60/282,981
: PRIOR FILING DATE: 2001-04-11
: PRIOR APPLICATION NUMBER: 60/283,656
: PRIOR FILING DATE: 2001-04-13
: PRIOR APPLICATION NUMBER: 60/309,247
: PRIOR FILING DATE: 2001-07-31
: PRIOR APPLICATION NUMBER: 60/311,754
: PRIOR FILING DATE: 2001-08-17
: PRIOR APPLICATION NUMBER: 60/313,331
: PRIOR FILING DATE: 2001-08-17
: NUMBER OF SEQ ID NOS: 260
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 40
: LENGTH: 914

```

; LIFE: PKI  
; ORGANISM: Homo sapiens

### Query Match

18.41; Score 876.5; DB 12; Length 914;



```
QY 163 TSPDSSDNTYF:FKDOFLQSSALPENRNLXGIGETHKSRLL---IPGETMTLWAD 219
Db 251 TFXTHSDSKY---GMSVGLDFSLP-GMEHVYIGIPEHAD-NLRKVKTEGGPPYRLYND 305
QY 220 IGS---ENPDVNIYGHSP-----FYMDVRGS-----KGNEEAGTTHGVLL--L 257
Db 306 VFOYELYNP-MALYGSVPVLLAHNPHRDLGIFWLNAAETWVDISSNTACKTLFGKWMOYL 364
QY 258 NSNG-----MDVKYEGHRTITNVIGGVIDLYVFGPSPFEMVMQYTELGRPAMPYWSG 313
Db 365 QSGSETPQTRWMSSET-----GIIDVELLGPISDVFRQYASLTGTQALPLFLSLG 417
QY 314 FHOCRYGKXVNSLEYVVDYKAGAPLEVMWTDIDYMDGYKDFTLDPVNFPEDKMQSFV 373
Db 418 YHOSRWYREADVLEVDQGFDDHNLPCDVIWLDIEHAGKRYFTWDPDRFFQPR--TML 475
QY 374 DTLHKGQKVLILDPGIGVDSSYGTNRGMEADVFIK-RNGEPYLGVMYKGVFPDFL 432
Db 476 ERLAKRKLVAIVDPHIKVDGSRVHEELRNGLYVKTRDGSYDEGCMWPGSAGYPDF 535
QY 433 NPAATAFWSNEIKM--FOETILPLDGLWDMNVEISNFIITSLSSGSLDDPPYKINNSGDK 490
Db 536 NPTWRANWAMFSDNYEGSAPNLFVWMDNPEVSF-----NGPEV----- 576
QY 491 RPINNTVPATSIHFGNISEYDAHNLGLLEAKATHQAVVDITG--KRPFILSRSTFVSS 548
Db 577 -----TMLKDAHYGWEHRDVHNIYGLVYHNTADGLRQSRSGMERPPVLARAFFAGS 630
QY 549 GKTAHTWGNAAKWEDLAYSIGILNFGILFGIPMYGADICGFSHTTTELCCRNIQLGA 608
Db 631 ORFGAVMTGNTAEWOHLKISIPMCLSLGLVGLSFCGADVGVGFKNPEPELLVRYQMGA 690
QY 609 FYFPAHDHSLGTAROEYLWDSVASS-ARKVLGLRMLLPHLYT-MYEAHVSGRPIARP 667
Db 691 YQPFPAHAHLDTGRREPWLPHQNDITRDALQYISLLPFWYLLYQAHREGIPWMP 750
QY 668 LFSFPQDTKYEIDSOFLGKSIMVSPALKQCAVADAYFPAGN--WFDLFNYSFAVG 725
Db 751 LWQYQPDVTTFTFIDQYLLIGDALLVHPVSDSAGHVQVYLPQCGEVVYDIOSYQ----- 805
QY 726 DSGH-----VRLDTPADHVNHVREGSIVAMQGEALTTREDARK-TPYOLLVASELENI 779
Db 806 ---KHGPGQLYLPVLUSSIPVFORGTIVPRMVRRSSECKMDPITLVALSPQGT 862
QY 780 SGELFLDDGLENRMAGGGRNDTLVKFRCYVTGKSVLRSEVNVNPEYASKMKSIGKYT 839
Db 863 QGELFLDDGHTFNQVT---RQEFLLRFS--FSGNTLV--SSADPEGHFEFTPIWIERV 915
QY 840 FVG 842
Db 916 IIG 918
```

## RESULT 1C

US-10-032-189-42

; Sequence 42, Application US/10032189

; Publication No. US20030170630A1

; GENERAL INFORMATION:

; APPLICANT: Alsbrook II, John P

; APPLICANT: Tchernov, Velizar T

; APPLICANT: Liu, Xiaohong

; APPLICANT: Zyrtex, Kimberly A

; APPLICANT: Zerhusen, Bryan D

; APPLICANT: Patturajan, Meera

; APPLICANT: Grosse, William M

; APPLICANT: Lepley, Denise M

; APPLICANT: Burgees, Catherine E

; APPLICANT: Shimkets, Richard A

; APPLICANT: Grosse, William M

; APPLICANT: Szekeres, Edward S

; APPLICANT: Vernet, Corine A.M.

; APPLICANT: Li, Li

; APPLICANT: Casman, Stacie J

```
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Gorman, Linda
; APPLICANT: Gangolli, Esha A
; APPLICANT: Fernandes, Elma R
; APPLICANT: Rieger, Daniel K
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Gunther, Erik
; APPLICANT: Millet, Isabelle
; APPLICANT: Sciore, Paul
; APPLICANT: Ellerman, Karen
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glennda
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-228
; CURRENT APPLICATION NUMBER: US/10/032,189
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,495
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/258,171
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/269,940
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 60/274,192
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/277,826
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/279,840
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/282,981
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/283,656
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/309,247
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/311,754
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/313,331
; NUMBER OF SEQ ID NOS: 260
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 42
; LENGTH: 914
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-032-189-42
```

Query Match 18.4%; Score 875.5; DB 12; Length 914;

Best Local Similarity 27.9%; Pred.No.3.8e-72;

Matches 256; Conservative 155; Mismatches 319; Indels 187; Gaps 34;

```
QY 84 RLRIITDSSQORWEIPEITVI PRAGNHSRRFSTEEDGGNSPENNFLOPSSDLVFTLHN 143
Db 83 RLKINETPLKPRFEVDPVLTSPKS---TVRLDSCSGDTGSL---ILADKGDL--KCHI 134
QY 144 T-TPFGFS-----VSRSSGDILF-----DTSPPSSDNTYFIEX 177
Db 135 TANPFKVDLVSEERWVISINSLGQLYFEHLQILHKORAAKENEETSVDTSQENQ---E 190
QY 178 D-----QFLOLSSALPEN-----RNLVYIGETHYKRSFRLI---PGETMTLW 216
Db 191 DLGLWEEKFGKFDVLIKANGPSSIGLDFSLHGFHLYGIPQHAIE-SHQLKNTGDDAYRL 249
QY 217 NADI-GSENPD-VNLYGSHPPFYMDVRGSKNGEAGTTHGVLLNSNG--MDVKYE----- 267
Db 250 NLDVYGVQIYDKMGVYGVYLL-----AHKLGRTIGIFNLNASETLVEINTEPAVEY 302
QY 268 -----GHR--ITYNVIGGVIDLIVYFAGSPFEMVMNQYTELIGRPAMPYWS 311
Db 303 TLTQMGVPAAKQKVGSRTHVHMWSESGLIDVFLLTGTPTSDVFKQYSHLTGTQAMPPLFS 362
QY 312 PGFHQCRYGYNVSDLEYVVDGYAKAGIPLEVAMTWTDIDYMDGYKDFTLDPVNFPEDKMQS 371
Db 363 LGYHQCRRWYDEQDVKAVDAGDFDEHDIPYDAMWMLDIEHTGKRYFTWQKRNFPNPKRMQ 422
```





Db 554 -----TWLXDAQHYGGWEHRDVHNIYGLYVHMATADGLRQSGMERPFVLARFAFGSQ 605  
QY 550 KYTAHTWTGMAAKWEDLAYSPGILNPLFGIPMWGADICGFSHDTTEELCRRWIQLGAF 609  
Db 609 RFGAVMVGENTAEMDHLKISIPXCLSLGLVGLFCGADVGFFKNPEPELLVWYQMGAY 668  
QY 610 YPFARDHSSLTAROEYLLWDSVASS-ARKVLGLRMRLPLHLYTLMYEAHVSGNPIARPL 668  
Db 669 QPFRAHAHLDTGREPWLLPSQHNDIIRDALGORISLLJFWWTLLYQAHRGPIVWRPL 728  
QY 669 FFSFPQDTKYEIDSQFLIGKISIMVSPALKQGAVDAYFPAGN--WFDLFNYSFAVGGD 726  
Db 729 WVOYPODVTTFNIDDOYLLGDALLVHPVSDSGAHGVQVYLPQGGEVWYDIQSYQ----- 782  
QY 727 SGKH-----VRLDTPADHVNHVHREGSIVAMQGEALTRDARK-TPYQLLVASRLNIS 780  
Db 783 --KHHGQOTLYPLVPTLSSIPVFORGGTIVRWRVRSSSECMKDDPTILFVALSPQGTQA 840  
QY 781 GELFLDDGENLRMGAGCGNDRWTLVKFCYVTKSVVLRSEVVNPEYASQMKWSIGKVT 840  
Db 841 GELFLDDGYTFNYQT-----RQSFLLRRFS--PSGNTLV--SSADPEGHFETPIWIERV 893  
QY 841 VG 842  
Db 894 IG 895

RESULT 12  
US-10-032-189-122  
; Sequence 122, Application US/10032189  
; Publication No. US20030170630A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsobrook II, John P  
; APPLICANT: Tchernev, Velizar T  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Zerhusen, Bryan D  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Grosse, William M  
; APPLICANT: Lepley, Denise M  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Grosse, William M  
; APPLICANT: Szekeres, Edward S  
; APPLICANT: Vernet, Corine A.M.  
; APPLICANT: Li, Li  
; APPLICANT: Casman, Stacie J  
; APPLICANT: Boldog, Ferenc L  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gangolli, Esha A  
; APPLICANT: Fernandes, Elma R  
; APPLICANT: Rieger, Daniel K  
; APPLICANT: Edinger, Shlomit R  
; APPLICANT: Gunther, Erik  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Sciore, Paul  
; APPLICANT: Eilerman, Karen  
; APPLICANT: MacDougall, John R  
; APPLICANT: Smithson, Glenda  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-228  
; CURRENT APPLICATION NUMBER: US/10/032,189  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 60/257,495  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/258,171  
; PRIOR FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: 60/269,940  
; PRIOR FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 60/274,192  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/277,826  
; PRIOR FILING DATE: 2001-03-22

; PRIOR APPLICATION NUMBER: 60/279,840  
; PRIOR FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/282,981  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: 60/283,656  
; PRIOR FILING DATE: 2001-04-13  
; PRIOR APPLICATION NUMBER: 60/309,247  
; PRIOR FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: 60/311,754  
; PRIOR FILING DATE: 2001-08-17  
; PRIOR APPLICATION NUMBER: 60/313,331  
; PRIOR FILING DATE: 2001-08-17  
; NUMBER OF SEQ ID NOS: 260  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 122  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-10-032-189-122

Query Match 18.1%; Score 862; DB 12; Length 944;  
Best Local Similarity 30.2%; Pred. No. 7.4e-71;  
Matches 254; Conservative 127; Mismatches 315; Indels 146; Gaps 33;

QY 52 VLTA---KLDLIKPSVYAPDIKSLNLHVSLETSERLRIRITDSSQQRWE---IPETVI 104  
Db 150 ILTARPFRLDLEDSL-----LLSVNARGLLLEFEHORAPRVSOQSKDPAEGDGAQPEET- 204  
QY 105 PRAGNHSRRFSTEEDGSGSPENNFLADPSS-DLVFTLHNTTTPFGFSVRRKSSGILFDT 163  
Db 205 PRDGO-----KPEETQCKAEKD---EPGAWEEFTKTH----- 233  
QY 164 SPDSSDSNTYFIFKDOFLQLSSALPENSNLYGICEHTKSRPL---IPGETMTLWNADI 220  
Db 234 -----SDSKPY---GPMVGLDFSLP-GMEVVGIPEDHAD-NLRKLVKTSGGEPYRLYNLDV 284  
QY 221 GS---ENPDVNLYGSHP-----FYMDVRGS-----KGNEEAGTTHGVLL--LN 258  
Db 285 FOYELYNP-MALYGSVPVLLAHPHRLDGLIFWLNAETWVDISSNTAGTKTLFGKMDYIQ 343  
QY 259 SNG---MDVKYEGHRIYTNVIGVIDLVYFAGSPSPMMQYTELIGRPAMPYWSFGF 314  
Db 344 GSGETPQTDVWRWSET-----GIIDVFLGLGPSISDVFRQYASLTGTQALPPLFLSLGY 396  
QY 315 HOCRYGYNVSDLEVYVDGYAKAGIPLVMWTDIDYMDGYKDFLDPVNFPEDKMQSFVD 374  
Db 397 HOSRWNYRDEADVLEVDQGFDDHNLPCDVIWLDIEHADGKRYFTWDPGRFPQPR--TWLE 454  
QY 375 TLHQNGQKYVLLIDFGIGVDSSYGTYNRGMEADVFIK-RNGEPYLGVEVMPGKVYFPDFLN 433  
Db 455 RLASKRRLVAIVDPHIKVDGSGYRVHEELRLNLGLVYKTRDGSYEGWCWPGSAGYPDFTN 514  
QY 434 PAAATFWSNEIKM--FOEILPLDGLWIDMNELSNFIITSPSSGSSLDPPYKINNNGDKR 491  
Db 515 PTMRANWANMFSDNYEGSAPNLFYVNDMNEPSVF-----NGPEV----- 554  
QY 492 PINNKTVPATSIHFGNISYDAHNLYGLLEAKATHQAVVDITG--KRPFILSRSTFVSSG 549  
Db 555 -----TMLKDAOHYGGWEHRDVHNIYGLYVHMATADGLRQSGMERPFVLARFAFGSQ 609  
QY 550 KYTAHTWTGMAAKWEDLAYSPGILNPLFGIPMWGADICGFSHDTTEELCRRWIQLGAF 609  
Db 610 RFGAVMVGENTAEMDHLKISIPXCLSLGLVGLFCGADVGFFKNPEPELLVWYQMGAY 669  
QY 610 YPFARDHSSLTAROEYLLWDSVASS-ARKVLGLRMRLPLHLYTLMYEAHVSGNPIARPL 668  
Db 670 QPFFRAHAHLDTGREPWLLPSQHNDIIRDALGORISLLJFWWTLLYQAHRGPIVWRPL 729  
QY 669 FFSFPQDTKYEIDSQFLIGKISIMVSPALKQGAVDAYFPAGN--WFDLFNYSFAVGGD 726  
Db 730 WVOYPODVTTFNIDDOYLLGDALLVHPVSDSGAHGVQVYLPQGGEVWYDIQSYQ----- 783  
QY 727 SGKH-----VRLDTPADHVNHVHREGSIVAMQGEALTRDARK-TPYQLLVASRLNIS 780

Db 784 --KHGPGQTLXPVTLSSIPVFGGCTIVPRWVRSSCKMKDDPITLFLVALSPQGTAA 841  
QY 781 GELFLDDGGLNLRGAGGGRDWTLYKFCYVTKGKSVLSEVWNPYAKKMSICKUTTF 840  
Db 842 GELFLDDGTYFNQYOT--RQEFLLRRFS--FSGNTLV--SSADPEGHFTPIWIERVYI 894  
QY 841 VG 842  
Db 895 IG 896

## RESULT 13

US-10-032-189-121  
; Sequence 121, Application US/10032189  
; Publication No. US20030170630A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsbrook II, John P  
; APPLICANT: Tchernev, Velizar T  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Zerhusen, Bryan D  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Grosse, William M  
; APPLICANT: Lepley, Denise M  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Grosse, William M  
; APPLICANT: Szekeres, Edward S  
; APPLICANT: Vernet, Corine A.M.  
; APPLICANT: Li, Li  
; APPLICANT: Casman, Stacie J  
; APPLICANT: Boldog, Ferenc L  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gangolli, Esha A  
; APPLICANT: Fernandes, Elma R  
; APPLICANT: Rieger, Daniel K  
; APPLICANT: Edinger, Shlomit R  
; APPLICANT: Gunther, Erik  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Sciote, Paul  
; APPLICANT: Ellerman, Karen  
; APPLICANT: MacDougall, John R  
; APPLICANT: Smithson, Glennda  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-228  
; CURRENT APPLICATION NUMBER: US/10/032,189  
; PRIOR FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 60/257,495  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/258,171  
; PRIOR FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: 60/269,940  
; PRIOR FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 60/274,192  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/277,826  
; PRIOR FILING DATE: 2001-03-22  
; PRIOR APPLICATION NUMBER: 60/279,840  
; PRIOR FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/282,961  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: 60/283,856  
; PRIOR FILING DATE: 2001-04-13  
; PRIOR APPLICATION NUMBER: 60/309,247  
; PRIOR FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: 60/311,754  
; PRIOR FILING DATE: 2001-08-17  
; PRIOR APPLICATION NUMBER: 60/313,331  
; PRIOR FILING DATE: 2001-08-17  
; NUMBER OF SEQ ID NOS: 260  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 121

; LENGTH: 966  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-10-032-189-121  
Query Match 17.9%; Score 851.5; DB 12; Length 966;  
Best Local Similarity 30.2%; Pred. No. 7.3e-70;  
Matches 242; Conservative 116; Mismatches 301; Indels 141; Gaps 30;  
QY 52 VLTAA--KLDLIKPSVYAPDINKSLNH--VSLETSELRIRITDSSQ-----QRWEIPET 102  
Db 150 ILTAQPFRLDLEDRSL-----LLSVNARGLMAFEHORAPRVPSFKVSLALGSGVMDKIK 205  
QY 103 VIPRAGNHSP-----RRFSTEEDCGNSPENNFADPSS-----DLVFTLH-NTTPFGFSV 151  
Db 206 LFSRQESKDPAGNCAQPEATPGDGKPKPETQKAEDKDEGAEETFKTHSDSKPYG--- 262  
QY 152 SRRSSGDLFDTPSDSSDNTYFIKQDFLOJSSALPENRSLNYGIGEHTKGSRFL---I 208  
Db 263 -----PTS-----VGLDFSLP-GMEHVYGIPEHAD-SRLKRVTE 294  
QY 209 PGETMTLWNADIGS---ENPDVNLVGS-----HPFYMDVR-----GSKNEEA 248  
Db 295 GGEPRYLNDVFOYELNRP-MALYGSVPVLLAHSFHRDLGIFPLNAAETWVDISNTAG 353  
QY 249 GTTHGVLL--LNSNG-----MDVKYEGHRITYNVIGVIDLYVPAGSPENVMNQYTELIG 302  
Db 354 KTLFGKMLDYLGSGETPQTDIRMSE-----SGIDVFLMLGSPVDFVFRQYASLTG 406  
QY 303 RPAPMYWSEFGHOCRYGYKXNSDLEYVDVGYAKAGIPIEVMWTDIDYMGYKDFLDPV 362  
Db 407 TQALPPLFSLSGTHQSRWNYRDEADVLEVDQGFDDHNMPCDVIWLDIEHADGKRYFTWDEP 466  
QY 363 NFPEDKMQSFVDTLHKNGQKYLILDPGIGVDSYGYTNRGMEADVFIK-RNGEYVLGEV 421  
Db 467 RFPQP--LNMLEHLASKRKLVAVDPHIXVDSGYRVHEELRNHGLYVTRDSDYEGWC 524  
QY 422 WPGKYVFPDFLNPAATFTMSNEIKM--FOEILPLDGLWIDMNELSNFIITSPSSGSLDD 479  
Db 525 WPGSASYPDFTNPRMRAWNSMFSFONYEGSAPNLVYMDMNEPSVF-----NGPEV-- 576  
QY 480 PPYKINNSGDKRFINNKTVPATSIHFGNISEYDAHNLGLLEAKATHQAVVDITG--KRP 537  
Db 577 -----TMLKDAVHYGGWEHRDIHNIYGLYVHMATADGLIQSGGIERP 619  
QY 538 FILSRSTFVSSGKYTAHWTGDNAAKWEDLAYSPGILNFGILGIPMWGADICGFSHDTTE 597  
Db 620 FVLSRAFFSGSQRFQFVAVWTGNTAEWDHLKISIPMCLSLALVGLSFCGADVGGFKNPEP 679  
QY 598 ELCRWITQLGAFYFPFADHSSLSGTARQELYMDSVASSA-RKVGLRLMLLLPHLYTLMYE 656  
Db 680 ELLVRYWYQMGAYQPFRAHAHLDTGRREPWLLASQYQDAIRDALFORYSLLPFWYTLFYQ 719  
QY 657 AHVSGNPIARPLFFSPQDTKTYEIDSQFLIGKISIMVSPALKOGAVADVAYFPAGN--WF 714  
Db 740 AKEGFPVMPRLWQVQPEDVNSTFIEDQFMLGDALLIHPVSDAGANGVQVYLPQGEVWY 799  
QY 715 DLFNYSFAVGSDSGKH-----VRLDTPADHVNVHVRSGSIVAMQGEALTRDARK-TPYQ 768  
Db 800 DIQSYQ-----XKHGPQTLVLPVTLSSIPVQFGGTTIVPRWVRSSDCMKDDPIT 851  
QY 769 LLVVASRLNISGELEFLDDG 788  
Db 852 LFVALSPQGTAAQGELEFLDDG 871

## RESULT 14

US-10-032-189-38  
; Sequence 38, Application US/10032189  
; Publication No. US20030170630A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsbrook II, John P  
; APPLICANT: Tchernev, Velizar T

APPLICANT: Liu, Xiaohong  
APPLICANT: Spytek, Kimberly A  
APPLICANT: Zehrusen, Bryan D  
APPLICANT: Fatturajan, Meera  
APPLICANT: Grosse, William M  
APPLICANT: Lepley, Denise M  
APPLICANT: Burgess, Catherine E  
APPLICANT: Shimkets, Richard A  
APPLICANT: Grosse, William M  
APPLICANT: Szekeres, Edward S  
APPLICANT: Vernet, Corine A.M.  
APPLICANT: Li, Li  
APPLICANT: Casman, Stacie J  
APPLICANT: Boldog, Ferenc L  
APPLICANT: Gorman, Linda  
APPLICANT: Gangolli, Esha A  
APPLICANT: Fernandes, Elma R  
APPLICANT: Rieger, Daniel K  
APPLICANT: Edinger, Shlomit R  
APPLICANT: Gunther, Erik  
APPLICANT: Millet, Isabelle  
APPLICANT: Sciore, Paul  
APPLICANT: Ellerman, Karen  
APPLICANT: MacDougall, John R  
APPLICANT: Smithson, Glennda  
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-28  
CURRENT APPLICATION NUMBER: US/10/032,189  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: 60/257,495  
PRIOR FILING DATE: 2000-12-21  
PRIOR APPLICATION NUMBER: 60/258,171  
PRIOR FILING DATE: 2000-12-20  
PRIOR APPLICATION NUMBER: 60/269,940  
PRIOR FILING DATE: 2001-02-20  
PRIOR APPLICATION NUMBER: 60/274,192  
PRIOR FILING DATE: 2001-03-08  
PRIOR APPLICATION NUMBER: 60/277,826  
PRIOR FILING DATE: 2001-03-22  
PRIOR APPLICATION NUMBER: 60/279,840  
PRIOR FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: 60/282,981  
PRIOR FILING DATE: 2001-04-11  
PRIOR APPLICATION NUMBER: 60/283,656  
PRIOR FILING DATE: 2001-04-13  
PRIOR APPLICATION NUMBER: 60/309,247  
PRIOR FILING DATE: 2001-07-31  
PRIOR APPLICATION NUMBER: 60/311,754  
PRIOR FILING DATE: 2001-08-17  
PRIOR APPLICATION NUMBER: 60/313,331  
PRIOR FILING DATE: 2001-08-17  
NUMBER OF SEQ ID NOS: 260  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 38  
LENGTH: 967  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-032-189-38

Query Match 17.6%; Score 837.5; DB 12; Length 967;  
Best Local Similarity 29.3%; Pred. No. 1.5e-68;  
Matches 24; Conservative 130; Mismatches 330; Indels 121; Gaps 28;  
QY 26 VLEESTV-----VGYVVRVGVDSNRQVLT---KLDLKPSVYAPDTKSLNLHV 77  
Db 121 VLSKSTVRLSCSGTGSLLADGKDLKCHITANPFKVDLVSEEVVI-SINSLG-OL 178  
QY 78 SLETSELRIRITDSQ-----QRWEIPETVPRAGNHSRRFSTEEDGGNSPENNFLADP 133  
Db 179 YFEHGRAPRVFSFDKVNLTLSIWDKIKNLSRQSGKDP---AEGDGAQPEETPRDGBK 234  
QY 134 SSDLVFTLHNTTFPGVSRRSSGDLFDTSPPSSSNNTYFFKQDFLOLSALPENRSN 193

Db 235 PEETQGAKEKDEPGAW-----ETFKTHSDSKPY---GPSSIGLDFSL-HGFEH 279  
QY 194 LYGIGHTKSRFLI-----PGETMLNADI-GSENDP-VNLYGSHPPYMDVRSGKNEEA 248  
Db 280 LYGIPQHAH-SHOLKNTGDDAYRLVNLVDYGYQIVDKMGIVGVPYLL-----AHKL 331  
QY 249 GTTHGVLLNSNGMVKYEGHRITYNVIG-----GVIDLYV 284  
Db 332 GRTIGIFWLNASETLVEINTEPAGIVIFGVPVSIYOSQDTPLTTHVHMSESGIIDLVL 391  
QY 285 FAGSPMVMNQYTELIGRPAPMPYMSFGHQCRGYKXVSDLEYVYVGVGAKAGIPLVEM 344  
Db 392 LTGTPSDVFKQYSHLTGTQAMPPLFSLGYHOCRWNYEEDQVKAVDAGDEHDIYDAM 451  
QY 345 WTDIDYMGYKDFTLDPVNFPEKMQSFVDTLHKNGQKYVLLIDPGIGVDSSVGTNRGM 404  
Db 452 WLDIEHTGKRYFTMDKNRFPNPKRMQ--ELLRSKKRKLVIISDPHIKIEPDYSVYKAK 509  
QY 405 EADVFIK-RNGEYVLGEVWPVKVYFDFLNPAAATFMSNEIKMFQELPLDG-----LW 457  
Db 510 DQGFVNQGEDEFGVCNFMKSYLDFTNPKVREWYSS--MFSS--NCGGTDILFLW 564  
QY 458 IDMNELSNFITSPLSSGSSLDPPYKINNSGDKRPINNKTVPATSIHFGNISYDAHLNY 517  
Db 565 NDMNEPSVF-----RGP--EQTQKNAIHGHNWEHRLHNY 599  
QY 518 GLLEAKATHQAVVDITGKRPFILSRSTFVSSGKYTAHWTGDNAAKWEGLAYSIPGILNFG 577  
Db 600 GFYMATAGLIKRSKGRPFVLTSPFAGSQKYGAVTGDNTAEWSNLKISIPMLLTLS 659  
QY 578 LFGIPMYGADICGFSHDTTBELCRRWILQALGAFYFPFARDHSSLSGTARQELYL-DSVASSA 636  
Db 660 ITGISFCGADIGFIGNPETELLVRWYQAGAYQFFRFGHATMTKREPWFGEHRLI 719  
QY 637 RKVLGLRMLLPHLYTLMYEAHVSNGNPAPLPSPQDTKTYEISQFLIGKSIWVSPA 696  
Db 720 REAIRERYGLPYWYSLFYHARVASQPMRPLWVEFDELKTFDMEDEYMLGSALLVHPV 779  
QY 697 LKQGAVAVDAVFFAGN---WFDLFNTYSFVAGDSGK-HVRLDTPADHVNHVHREGSIVAM 752  
Db 780 TEPKATTVDFLPGSNEVWYDYKTFHWEGGCTVKIPVLLOIP-----VFQGGSVIPI 834  
QY 753 QGE-ALTRDARKTPYQLLVVASRLENIS-GEFLDDGENLR 792  
Db 835 KTTVGKSTGWTGWTSSYGLRVALSTLQSSVGVGLYLDGDSHFQ 876  
RESULT 15  
US-10-032-189-124  
; Sequence 124, Application US/10032189  
; Publication No. US20030170630A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsobrook II, John P  
; APPLICANT: Tchernev, Velizar T  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Zehrusen, Bryan D  
; APPLICANT: Fatturajan, Meera  
; APPLICANT: Grosse, William M  
; APPLICANT: Lepley, Denise M  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Grosse, William M  
; APPLICANT: Szekeres, Edward S  
; APPLICANT: Vernet, Corine A.M.  
; APPLICANT: Li, Li  
; APPLICANT: Casman, Stacie J  
; APPLICANT: Boldog, Ferenc L  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gangolli, Esha A  
; APPLICANT: Fernandes, Elma R  
; APPLICANT: Rieger, Daniel K  
; APPLICANT: Edinger, Shlomit R

APPLICANT: Gunther, Erik  
APPLICANT: Millet, Isabelle  
APPLICANT: Scioren, Paul  
APPLICANT: Ellettman, Karen  
APPLICANT: MacDougall, John R  
APPLICANT: Smithson, Glenn  
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-228  
CURRENT APPLICATION NUMBER: US/10/032,189  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: 60/257,495  
PRIOR FILING DATE: 2000-12-21  
PRIOR APPLICATION NUMBER: 60/258,171  
PRIOR FILING DATE: 2000-12-20  
PRIOR APPLICATION NUMBER: 60/269,940  
PRIOR FILING DATE: 2001-02-20  
PRIOR APPLICATION NUMBER: 60/274,192  
PRIOR FILING DATE: 2001-03-08  
PRIOR APPLICATION NUMBER: 60/277,826  
PRIOR FILING DATE: 2001-03-22  
PRIOR APPLICATION NUMBER: 60/279,840  
PRIOR FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: 60/282,981  
PRIOR FILING DATE: 2001-04-11  
PRIOR APPLICATION NUMBER: 60/283,656  
PRIOR FILING DATE: 2001-04-13  
PRIOR APPLICATION NUMBER: 60/309,247  
PRIOR FILING DATE: 2001-07-31  
PRIOR APPLICATION NUMBER: 60/311,754  
PRIOR FILING DATE: 2001-08-17  
PRIOR APPLICATION NUMBER: 60/313,331  
PRIOR FILING DATE: 2001-08-17  
NUMBER OF SEQ ID NOS: 260  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 124  
LENGTH: 944  
TYPE: PRT  
ORGANISM: Sus scrofa  
US-10-032-189-124

Query Match 17.5%; Score 832; DB 12; Length 944;  
Best Local Similarity 29.8%; Pred. No. 4.7e-68;  
Matches 242; Conservative 113; Mismatches 314; Indels 144; Gaps 28;  
QY 52 VLTA---KLDLTKPSVYAPDIKSLNLHVSLETSERLRIRITDSSQQRWEIPEVPIPRAG 108  
DB 150 ILTARPFRLDLEDRSL---LLSYNARGLLNFEHQAPRVSGSKD----- 192  
QY 109 NHPRRFSTEDCGNSPENNFLADPSSDLVTLHNTTFFGFSVSRSSGDLFDTSDDSS 168  
DB 193 -----PAEGGAQPEAPGDGKPEIQGAKEDEFGAME-----ETFKTHS 234  
QY 169 DSNTYFIKDFQLQSSALPENRSLNYGIBGHTKRSFRL---IPGETMTLWADIGS--- 222  
DB 235 DSKPY---GPTSVGLDFSLP-GMEHVYGIPEHAD-SLRLLKVTGGDPVRLYNLDVFOYEL 289  
QY 223 ENPDVNLVGSHP-----FYMDVRGS-----KGNEEAGTTHGVLL--LNSNG-- 261  
DB 290 YNP-MALYGSVFLLAHSPHRDLGIFLWLNAAETWVDISSNTAGTKLFGKMLDYLGSGSET 348  
QY 262 --MDVKYEGHRTYTVIGGVLDVYFAGFSPMVYVNYQYTELIGRPAPMPYWSFGHOCRY 319  
DB 349 PCTDVRWNE-----SGILDVFLLLGFSVDFVFRQYASLTGTQALPPLFSLGYHOSRW 401  
QY 320 GYKNVSDLYVVDGYAKAGIPLEVWMTDIDYMDGYKDFTLDPVNFPEPKMQSFVDTLHKR 379  
DB 402 NYRDEADVLEVNQGDHNLPCDFIWLDEHADGKRYFTWDPSPRFPOR--TMLEHLASK 459  
QY 380 GOKYVLLIDPGICLVDSYGYTVNRGHEADVFIK-RNGEYPLGEVNRGKYVPEDFLNPAAT 438  
DB 460 RRLKVAIVDPHIKVDSSYRVHEELQNLGLYVKTRDGSDEYGCWFGAASYPDFTNPKORA 519  
QY 439 FWSNEIKMFQELPLDG-----LWIDMNLNFTITSPSSGSLDPPPYKINNNGDKRP 492

DB 520 WYADMPR-FEN---YEGSSSNLYVWDMNEPSVF-----NGREV----- 554  
QY 493 INNKTVPATSIHFGNISEYDAHNLGLLEAKATHQAVVDITG--KRPFILSRSTTFVSSGK 550  
DB 555 ---TMLKDAQHYGGWEHRDLHNIYGFVHMATADGLVLRSGGVVERPVLRSRAFFAGSOR 61C  
QY 551 YTAHWTCGNAKMEDLAYSIPIILNFGILFGIPMVGADICGFSHDTTEELCERWIOLGAFY 610  
DB 611 FGAVMTGUNTAEWDHLKISIPMCUSLGLVGVSPCGADVGGFFKNPEPELVRYWQMGAYQ 670  
QY 611 PFARDHSSLSGTARQELYLWDS-VASSARKVLGLMRLLPHLYTILMYEAHVSGNPIARPLF 669  
DB 671 PFFRAHAHLDTGRREPWLLPTQYQDMIRDALGQRYSLLPFWYTLFYQAHREGVPMRALW 730  
QY 670 PSFPQDTKYEIDSQFLIGKSIMVSPALKQGAVAVDAYFPAGN--WEDLFNYSFVAVGDS 727  
DB 731 VHYPODVTTFISIDDEFLLDALLVHPVTDSEAHGVQVYLFQGGSEVWYDVHSHYOKYHGPQT 790  
QY 728 GKRVRLDTPADHVNHVHVRGSIAMQGEALTTRDARK-TPYQLLVVASRLENISGEFLFD 786  
DB 791 ---LYLPVTLSSIPVFQGGTIVPRWVRSSDCMKDDPITLFLVALSPQGTAGELFLD 847  
QY 787 DGENLRMGAGGGRDWTLVKFRCTVVTGKSVVLR 819  
DB 848 DGHTFN-----YQTGHEFLLR 863  
Search completed: October 27, 2003, 10:38:19  
JCS time : 112.36 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 ; Search time 16.3042 Seconds  
(without alignments)  
2340.767 Million cell updates/sec

Title: US-10-043-418-4

Perfect score: 4751

Sequence: 1 MSSLHWFPNIFIVVVFFSL.....EVSLSLJGKKFEXELR.T 902

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/2/iaa/5A COMB.pcp.\*
- 2: /cgn2\_6/ptodata/2/iaa/5B COMB.pcp.\*
- 3: /cgn2\_6/ptodata/2/iaa/6A COMB.pcp.\*
- 4: /cgn2\_6/ptodata/2/iaa/6B COMB.pcp.\*
- 5: /cgn2\_6/ptodata/2/iaa/PCBUS.COMB.pcp.\*
- 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pcp.\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2451	51.6	877	1	US-08-430-925A-4
2	858.5	18.1	914	4	US-09-437-054A-8
3	833	17.5	919	4	US-09-437-054A-17
4	713.5	15.0	693	4	US-09-376-343-2
5	668	14.1	938	4	US-08-897-843A-1
6	494.5	10.4	1070	2	US-08-633-770A-2
7	479.5	10.1	1066	2	US-08-633-770A-1
8	366	7.7	1091	3	US-08-633-768A-2
9	357.5	7.5	1088	3	US-08-633-768A-1
10	356.5	7.5	1092	4	US-09-275-608-3
11	333.5	7.0	570	4	US-09-275-608-4
12	190	4.0	390	4	US-09-634-238-331
13	185	3.9	150	4	US-09-437-054A-10
14	131.5	2.8	906	4	US-09-367-895-41
15	126	2.7	1416	4	US-09-071-035-404
16	126	2.7	1448	4	US-09-071-035-402
17	117	2.5	3696	4	US-09-134-001C-5080
18	116.5	2.5	1245	4	US-09-252-991A-30935
19	116	2.4	1313	4	US-09-071-035-450
20	116	2.4	1313	4	US-09-071-035-454
21	115	2.4	327	4	US-09-134-001C-4937
22	114	2.4	1752	4	US-09-556-877-180
23	114	2.4	1752	4	US-09-620-412C-180
24	114	2.4	1752	4	US-09-598-419-180
25	114	2.4	1848	3	US-08-236-791-6
26	114	2.4	1848	5	PCT-US95-10661A-6
27	113.5	2.4	2504	4	US-08-851-567B-12

28	113	2.4	793	3	US-08-433-522A-10	Sequence 10, Appl
29	113	2.4	793	3	US-09-135-166-10	Sequence 10, Appl
30	113	2.4	793	3	US-08-942-046-10	Sequence 10, Appl
31	112.5	2.4	1337	3	US-08-854-585-2	Sequence 2, Appl
32	112.5	2.4	1337	4	US-09-447-533-2	Sequence 2, Appl
33	112.5	2.4	1337	5	PCT-US95-0512-2	Sequence 2, Appl
34	111	2.3	1702	3	US-08-296-791-5	Sequence 5, Appl
35	111	2.3	1702	5	PCT-US95-10661A-5	Sequence 5, Appl
36	110.5	2.3	262	1	US-08-332-828C-37	Sequence 37, Appl
37	110.5	2.3	262	3	US-09-330-945-37	Sequence 3, Appl
38	110.5	2.3	943	4	US-09-397-885-5	Sequence 5, Appl
39	110.5	2.3	943	4	US-09-969-362-5	Sequence 5, Appl
40	110.5	2.3	1577	2	US-08-793-824-2	Sequence 2, Appl
41	109.5	2.3	1612	1	US-08-169-927-2	Sequence 2, Appl
42	107.5	2.3	742	4	US-09-107-532A-4996	Sequence 4996, Ap
43	107.5	2.3	903	1	US-08-750-532-1	Sequence 1, Appl
44	107.5	2.3	1398	1	US-08-750-532-9	Sequence 9, Appl
45	107.5	2.3	1398	3	US-08-894-818B-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1  
US-08-430-925A-4  
; Sequence 4, Application US/08430925A  
; Patent No. 5763252  
; GENERAL INFORMATION:  
; APPLICANT: Skadsen, Ronald W  
; APPLICANT: Tibbot, Brian K  
; TITLE OF INVENTION: Cloned Alpha-Glucosidase from Barley  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: 1 South Pinckney Street  
; CITY: Madison  
; STATE: WI  
; COUNTRY: US  
; ZIP: 53703  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/430,925A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J  
; REGISTRATION NUMBER: 27,386  
; REFERENCE/DOCKET NUMBER: 960296.93171  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 608-251-5000  
; TELEFAX: 608-251-9166  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 877 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-430-925A-4

Query Match 51.6%; Score 2451; DB 1; Length 877;

Best Local Similarity 55.2%; Pred. No. 9.6e-225;  
Matches 471; Conservative 136; Mismatches 191; Indels 56; Gaps 13;

QY	69	DIKSLNHLVSLTGERLIRITDSQQRWIPETVIPRAG-----NHSPPRFSTEEDCGN	123
DB	59	DVORLAVVASLETDSRLVRITDADHPWEVQDIIRPAPGDLVHDAPPASSAPLOG--	116
QY	124	SPENNFLADPSSDLVFTLHNTTPFGFSVSRSSGDLFDTPSPDSSDNTYFIKDDQFIQL	183

```

Db 117 ----RVLSFAGSDVLVTH-ASPFRTVSRSTGTLFTAPG-----LVFRDKYLEV 164
Qy 184 SSALPENRSLNYGIGEHKTRFRILIPGETMTLWADIGSENPDVNLXGSHPEYMDVRGSK 243
Db 165 TSALPAGRASLYGGEHTKSFRLRHNDSTFTLWADIGASTVDVNLXGSHPEYMDVRA-- 222
Qy 244 GNEEAGTHGVLLNSNGMDVYKGEHRTIYVIGVIDLYVFAGSPSEMVNNOYTELIGR 303
Db 223 ----PGTAHGVLLSSNGMDVLYGSSVYVYKVGVLDFYFAGPNPLAVVDQYQLIAR 278
Qy 304 PAMPYMSFGFHQCRYGKYNVDLEYVDGAKAGIPLVNMWTDIDMDGDKBTLDPVN 363
Db 279 PAMPYMSFGFHQCRYGKYNVDLEYVDGAKAGIPLVNMWTDIDMDGDKBTLDPVN 338
Qy 364 FPEKMQSFVTLHKNGQKYLILDPGIGV--DSSYGTYNRGMADVFTKRCNEPYLGE 420
Db 339 FTAALREFVDRHLHNAQKYLILDPGIRVDPIDATYGTFRGQQDIFLKRNTNFVN 398
Qy 421 VMPGVYFPDFLNPAAATFWSNEIKMFOEILPLDGLWIDMNLNSFIT--SPLSSGSSLD 479
Db 399 VMPGVYFPDFLNPAAATFWSNEIKMFOEILPLDGLWIDMNLNSFIT--SPLSSGSSLD 455
Qy 480 PPKYKINSGDEKPPINNKTVPATSIHFGNISEYDAHNLGLLEAKATHQAVVDITGKRPFI 539
Db 456 PPKYKINSGDEKPPINNKTVPATSIHFGNISEYDAHNLGLLEAKATHQAVVDITGKRPFI 515
Qy 540 LSRSTFVSSGKYTAHWTGDNAAKWEDLAYSIPIGLFIPMWGADICGFSHDTTEEL 599
Db 516 LSRSTFVSSGKYTAHWTGDNAAKWEDLAYSIPIGLFIPMWGADICGFSHDTTEEL 575
Qy 600 CRRWIOGAFYPPARDHSSLGTAEOELYLMDSVASSARKVGLRMLRLLPLHYLTYLMEAHV 659
Db 576 CRRWIOGAFYPPARDHSSLGTAEOELYLMDSVASSARKVGLRMLRLLPLHYLTYLMEAHV 635
Qy 660 SGNPIARLPFSFPDQTKTYEIDSCFLGKIMVSPALKOGAVADAYEPAGWFDLPNY 719
Db 636 TGAPIARLPFSFPDQTKTYEIDSCFLGKIMVSPALKOGAVADAYEPAGWFDLPNY 695
Qy 720 SFAYGDSGKIVRLDTPADHNVHVRGSIIVAMQGEALTTDRDKTPYQLLVWASRLNI 779
Db 696 SLAVATRTGKHVRLPAPADTVNVHLTGTLPLQSSALTTSRABRTAFHLLVALAEGTA 755
Qy 780 SGELFLDGENLNRMGAGGNDWTLVPRCVYTGK--SVLRSRVNPEYASKMWSIGK 837
Db 756 SGYLFDDGDSPEYGR---RSDWSMVRFNKIPKNGKGAIKVSEVHNSVAQSRRLVISK 812
Qy 838 VTFVGFENVENVK--TYEVRTSERLRSRISLIKTVDNDPRLS-----VEVSKLS 886
Db 813 VLMGHRSPAPPKLITVHNSAE-----VEASSAGTRYONAGGLGGVAHIGLS 862
Qy 889 LLVGKGFEMRLRLT 902
Db 863 LVVGEFEELKAMS 876

RESULT 2
US-09-437-054A-8
; Sequence 8, Application US/09437054A
; Patent No. 6316698
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Kinney, Anthony J.
; TITLE OF INVENTION: Plant Alpha-Glucosidase II Homologs
; FILE REFERENCE: BB1273 US NA
; CURRENT APPLICATION NUMBER: US/09/437,054A
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: 60/107,909
; PRIOR FILING DATE: 1998-No. 6316698ember-10
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
; LENGTH: 914
; TYPE: PRT

```

```

; ORGANISM: Glycine max
US-09-437-054A-8

Query Match 18.1%; Score 858.5; DB 4; Length 914;
Best Local Similarity 30.1%; Pred. No. 1.6e-72;
Matches 244; Conservative 133; Mismatches 315; Indels 119; Gaps 29;

Qy 53 LTAKDLIKPSSVYAPDIKSLNLHYLSLETSERLRITDSS-----SQORWEIPEVI--- 104
Db 63 LTAKE---TPKHDSOSETKPLLLTUSVYQGRGLRLKIDEDPSLPKRFEPDVIDSEF 119
Qy 105 PRAGNHSPPRFSTEEDGGNSPENNLADPSSDLVFTLHNTTTPFGFSVSRSSGD----- 158
Db 120 PSTKLWLP-KISSVENGLSS--SVYLSDGHSVL-----RHPDFELFIRDDSGDRAVISLN 172
Qy 159 --ILPD--TSPDSDSNITYFIKQOFOQLSSALPENRNL-----YIGBHT 201
Db 173 SHDLDFEOLKHKSEDDN---WEEQFRSHTRDRRYPGQSIISFDVSFYGADRVYGIPIRA 228
Qy 202 KRSFLIP-----GETMTLWADIGS--ENPDVNLXGSHPEYMDVVRGSKGNEAGTT 251
Db 229 -ASLAKPTRGNVDESEPYRFLNDVPEYIHDSPFGLYGSIPFMVSHGKARGSS----- 282
Qy 252 HGVLNLNSNGM--DVYKEG-----HRTIT--YNVIGVIDLYVFAGSPSEMVNMQ 296
Db 283 -GFFWLNAABEQIDVLAFCWDAESGIALPSHRIDTFWMSAGVVDATFIFGNPKDVLQ 341
Qy 297 YTELGRPAMPYMSFGFHQCRYGKYNVDLEYVDGAKAGIPLVNMWTDIDMDGDK 356
Db 342 YTAVTGTAMPQLFIAHYHQCERNYRDEBDBHVDSEKFEDELIDYDVLWLDIEHTDGKRY 401
Qy 357 FTLDPNVPEE-DKMGOSFVDTLHKNGQKYLILDPGIVDSSYGTYNRGMADVFTK-RNG 414
Db 402 FTWDRALFHPPEMQR---KLAKGRHMTIIVDPHILKDENFHLHKEASOKYIYVDASG 458
Qy 415 EYELGEMVPGKYFPDFLNPAAATFWSNEI--KMFQEILPLDGLWIDMNLNSFITSPLS 472
Db 459 NDFDQWCMFPGSSSYPTLNPETRSWMAKFSQSYEGSTPSLYIWNDMNEPSVF----- 512
Qy 473 SGSSLDPPYKINSGDEKPPINNKTVPATSIHFGNISEYDAHNLGLLEAKATHQAVVDI 532
Db 513 NGPEV-----TMPRDVTHYGVHREHLNAYGYVPHMATANGLLK- 552
Qy 533 TGK---RPFILSRSTFVSSGKYTAHWTGDNAAKWEDLAYSIPIGLFIPMWGADIC 589
Db 553 RGEGRDNPVLSRALPAGSORVGAWVTGDNADHRLVSIIPWVLTGLTGNSPFGADIG 612
Qy 590 GFSHDTTEELCRRWIOGAFYPPARDHSSLGTAEOELYLW--DSVASSARKVGLRMLRLLP 648
Db 613 GFFGNPEPELLVRYQLGAYYPPFRAHAHHDTPKRREPWLFGERNTELIKDAIHRYALLP 672
Qy 649 HLYTLMYEAHVSGNPIARLPFSFPDQTKTYEIDSCFLGKIMVSPALKOGAVADAYE 708
Db 673 YPYTLFREANTTGVPRPLPMMEFFSDEATNSDTEFMVGSILVGGIYTERAKHASVYL 732
Qy 709 PA-GNMFEDLFNYSFAYGDSGKGVRLDTPADHNVHVRGSIIVAMQGE-ALTTRDKRTP 766
Db 733 PGKQSWYDLRTCAVYKGVVTHK---LEVTESIPAFQAGTIIARKDRFRSSSTQWANDP 789
Qy 767 YQLLVVASLENIISGELFDGGENLRMGAGG 797
Db 790 YTLVVALNSSQAAGELYIDGSSFNFLQGG 820

RESULT 3
US-09-437-054A-17
; Sequence 17, Application US/09437054A
; Patent No. 6316698
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Kinney, Anthony J.
; TITLE OF INVENTION: Plant Alpha-Glucosidase II Homologs
; FILE REFERENCE: BB1273 US NA

```

; CURRENT APPLICATION NUMBER: US/09/437,054A  
; CURRENT FILING DATE: 2001-05-14  
; PRIOR APPLICATION NUMBER: 60/107,909  
; PRIOR FILING DATE: 1998-No. 6316698ember-10  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 17  
; LENGTH: 919  
; TYPE: PRT  
; ORGANISM: Solanum tuberosum  
US-09-437-054A-17

Query Match 17.5%; Score 833; DB 4; Length 919;  
Best Local Similarity 28.7%; Pred. No. 4.4e-70;  
Matches 240; Conservative 128; Mismatches 298; Indels 170; Gaps 30;  
QY 71 KSLNHLVSLTSEIRLRITDSS-----QORWEIETVPIPRAGNHSPPRFTSEEDGNS- 124  
DB 80 KPLVLTLVSQGVQVMVKIDEDONLPPKKRFEVPEVI-----EEDFLNTK 125  
QY 125 -----PENNFLADSPSSDLVFTL-----HNTTFFGFSVSRSSGD---ILFDTSPD 166  
DB 126 LMLTRVKEQIDGVSSFSFVYLSDCYEGVLRHDP-----FEVFARESGKXVL----- 175  
QY 167 SSSDNTYFIP-----KQFLQLUSSALPENRNL-----YGEIHTK 202  
DB 176 SINSNGLFDEQLREKKEGDWEEKFRSHDTRPYGPQSISFDVSGFYGADFVYGIPEHA- 234  
QY 203 RSRLIP-----GETMTLWADIGS-----ENPDVNLVGHSPFMDVGRGSGNEEACT 250  
DB 235 TSFALKPTKGNVVEEYSEPRYLFNLDFEYLBHESP--FGLYGSIPFMISHGKARGSS--- 289  
QY 251 THGVLLLLNSGMDVKYEG-----HRI--TYNVIGGVIDLVVFGAGSPPE 291  
DB 290 --GFFWLNAAEQIDVLGSGWNSDESKIMLPDSKHRIOTLWMSGVDVTFEFGPEK 347  
QY 292 MYXNQVTELLIGRAPMYPMSFGHCQRYGKXVNSDLEYYVGVYAKAGIPLEVMTWTDIDYM 351  
DB 348 DVVRQYVTSVTRPSMPQLFATAYHOCRWYRDEEDVYVNSKFDHDIPIYDLVWLKDIEHT 407  
QY 352 DGYKFTLPVNVFPB--DKMQSFVDTLHKNGQKYLILDPGIGVDSYGYNGMGEADVFI 410  
DB 408 DGKKYFTMDRVVFPNPEEQK-----KLAAGRHWMTIVDPHIKDESYHTPKZALEKGYVY 464  
QY 411 K-RNGEPYLGEMVPGVYFPDFLNPAATFWSNEIKMFOELLPLOG-----LWIDM 460  
DB 465 KDATGXDYGWCPGSSSYTDLNPEIKSWSDKFS-----LSYVGSTKYLYIWNDM 517  
QY 461 NELSNTFPLSSGSLDDPPYKINNSGDKRPINNTVPATSIHFNGNISEYDAHNLGLL 520  
DB 518 NEPSVF-----NGPEV-----TMPRDALHGGVHEHRLHNSGYV 552  
QY 521 EAKATHQAVVDI--TGK-RPILSRSTFVSSGKYTAHTGDNAAKMEDLAYSIPGILNFC 578  
DB 553 FHNGTSDGLKXGDKDRPVLARAFAGSQRYGATWGDNTAEHLEHRSVPMVLTSI 612  
QY 579 FGIPMYGADICGSHDTEBELCRMTQLGAFYPPFARDHSSSLGTAROLYLW--DSVASSAR 637  
DB 613 SGIVPSGADVGFPGNPDTELLVRVYQVGYAYFFFGHAHHTDKREPMLFGERNTQLMR 672  
QY 638 KVLGLRMLLPHLYTLMYEARVSGNPIARLPFSFPQDTKTYEIDISQFLIGKSIWVSPAL 697  
DB 673 EATHVRMYLPPYPTLFRANSSTGTPVARPLWMEFFGDEKSFSDNEAFVWGNGLLVQGVY 732  
QY 698 KQGAVADAYFPA--GNWFDLFNYSFAVGDSGKHVRLDTPADHVNHVHREGSIVAMQGE- 755  
DB 733 TEPKHKVSVTLPGESWYDLRSASAYNGGHTHKY---EVSSEDSIPSFQAGTIIIRKDL 789  
QY 756 ALTRDARKTPYOLLVAGRLNENISGELFLDGCENLRMGAGGNRDMOTLVKFCVY 811  
DB 790 RRSSTQWENDPYTLVIALNSSKAAEGELYIDGKSYEFKGA-----FILKWEAVI 840

RESULT 4  
US-09-376-343-2  
; Sequence 2, Application US/09376343  
; Patent No. 6506592  
; GENERAL INFORMATION:  
; APPLICANT: Blum, Paul H.  
; TITLE OF INVENTION: Hyperthermophilic Alpha-Glucosidase Gene and Its Use  
; FILE REFERENCE: N1231-200  
; CURRENT APPLICATION NUMBER: US/09/376,343  
; CURRENT FILING DATE: 1999-08-18  
; EARLIER APPLICATION NUMBER: 60/096,860  
; EARLIER FILING DATE: 1998-08-18  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 693  
; TYPE: PRT  
; ORGANISM: Sulfolobus solfataricus  
US-09-376-343-2  
Query Match 15.0%; Score 713.5; DB 4; Length 693;  
Best Local Similarity 28.3%; Pred. No. 7e-59;  
Matches 182; Conservative 127; Mismatches 260; Indels 75; Gaps 23;  
QY 191 RSNLYGIGEHKRSFRL-IFGETMTLWADIGS-----ENPDVNLVGHSPFMDVGRGSKN 245  
DB 61 KEHIIGLGE---KAFELDRKRKRYVMYVNDAGAYKQYQDP---LYVSIPLFISVKD--- 110  
QY 246 EAGTTHGVLLNSNGM--DVKYEGH-RITYNVIGGVIDLVVFGAGSPPEMVMQYTELIG 302  
DB 111 ---GVATGYFPNSASKVIFDVGLEEDYDKVITVTPEDSVFVIEGPRIEDVLEKYTELGT 167  
QY 303 RPAPMPYMSFGHCQRYGKXVNSDLEYYVGVYAKAGIPLEVMTWTDIDYMGYKDFTLDPV 362  
DB 168 KPFLPPMWAQYMLSRYSYYPQDKAVVELVDIMQXGEFRVAGVFLDIHMYMSYKLTWHPY 227  
QY 363 NFPEDKMQSFVDTLHKNGQKYLILDPGIGVDSYGYNGMGEADVFKNGEPYLGEMV 422  
DB 228 RFPEPK--KLIDELHKNVKLITIVDHGIRVDQNSPFLSQMGKFCIE--SGELFVGKMW 284  
QY 423 PKYVYFPDFLNPAATFWSNEIKMFOELP--LDGLWIDMNLNSNFTITSLSSGSLDDP 480  
DB 285 PGTTYVPDFREDTREWAG---LISEWLSQGVDDGIWLDWNEPTDF--SRAIETRDVLSSL 340  
QY 481 PYKINNSGDKRPINNTVPATSIHF---GNISEYDAHNLVGLLEAKATHQAVVDITGKR 537  
DB 341 PVQFR---DRLVT--TFPDNVVHYLRGKVKHKKVKNAYFLYEMATFKGFRSHRNEI 395  
QY 538 FILSRSTFVSSGKYTAHTGDNAAKMEDLAYSIPGILNFCGIPMYGADICGFSH---- 593  
DB 396 FILSRAGVAGIQRYAFIWTGNTPSWDDLKLQLVLGLSISGVFVFGCDIGGQGRNFA 455  
QY 594 --DTEELCRWIOLGAFYPPFARDHSSL-GTAROLYLWDSVASSARKVLGLRMLLPHL 650  
DB 456 EIDMSMOLLVKYALALAFFPFRSHKATDGDITDTPVLPDYTKKVKKEIVELRYKFLPYI 515  
QY 651 YTLMYEARVSGNPIARLPFSFPQDTKTYEIDISQFLIGKSIWVSPALKQGAVADAYFPA 710  
DB 516 YSLALEASEKHVPVIRPLFYEFQDDDDMYRIEDYMWGVKYLAYPIVSKESRL-VTLPR 574  
QY 711 GNWFDLFNYSFAVGDSGKHVRLDTPADHVNHVHREGSIVAMQGEAL----- 757  
DB 575 GKYNYNW-----GEIINGKSVVKST--HELPIYLRGSIIPLEGDELIVYGETSFY 626  
QY 758 ----TTRDARKTPYOLLVVASRL-----ENISGELFLDGCENLRM 793  
DB 627 DNAEITSSNEIKFRSRIYVSKLTIITSEKPVSKIIIVDDSKETIQV 670

RESULT 5  
US-08-897-843A-1  
; Sequence 1, Application US/08897843A  
; Patent No. 6514493







Db 610 ---KXGNFLGSGVAGYRPAFLWTGDNASWFWKISVQVLSGLNGVCIAGSDT 665  
Qy 589 CGF-----SHDTTELC-----RRMQLGAFY2P-FARDH 616  
Db 666 GGFEPYDANGVEEKYCSPELLIRWYTGSELLPWLNRHYVKKRKMFPQEPYSPKHELEH 725  
Qy 617 SSLGTARQELVLDMSVASSARKVGLRWRLPLHLYTLMYEAHVSGNPIARPLFFSPQDT 676  
Db 726 PELA---DQAWLYKSVLEICRYVELYSLQLSDYCFMDFVQVGGXPIITRSMILLTDECT 782  
Qy 677 KTYE-----IDSQFLIGKINSVSPALKQGVAV-----DAYFPA-GNMF-----7-4  
Db 783 TFFNESQKFLDNQYMGAGDILVAIPILHSRKEIPGENDRVPLVPLVHTWYPSNLRPMDDQGV 842  
Qy 715 -----DLFNYSFVAGGDSGKHVRLDTPADH-----VNVHREGSIVAMQGEALTT 759  
Db 843 ALGNPVEGGSVINYT-----ARIVAPEDYNLFHSVVPVYVREGALIPQIEVRQWT 892  
Qy 760 RDARKTPVQLLVASRLNISGELFLDDGENLRMGAGGGRNDWTLVKFCY 810  
Db 893 -----GQGGANR-----IKENY 905

## RESULT 8

US-08-633-768A-2  
; Sequence 2, Application US/08633768A  
; Patent No. 6013504  
; GENERAL INFORMATION:  
; APPLICANT: YU, SHUKUN  
; APPLICANT: BOUSEN, KIRSTEN  
; APPLICANT: KRAH, KIRSTEN  
; APPLICANT: BOJKO, MAJA  
; APPLICANT: NIELSEN, JOHN  
; APPLICANT: MARCUSSEN, JAN  
; TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM  
; TITLE OF INVENTION: A FUNGUS INFECTED ALGAE, ITS PURIFICATION  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson & Bear  
; STREET: 620 Newport Center Drive 16th Floor  
; CITY: Newport Beach  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/633,768A  
; FILING DATE: 02-JUL-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 9321301.5  
; FILING DATE: 15-OCT-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Altman, Daniel E  
; REGISTRATION NUMBER: 34,115  
; REFERENCE/DOCKET NUMBER: DYOUT7.001APC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 714-760-0404  
; TELEFAX: 714-760-9502  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1091 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-633-768A-2

Query Match 7-7%; Score 366; DB 3; Length 1091;  
Best Local Similarity 22.2%; Pred. No. 2.6e-25;  
Matches 199; Conservative 133; Mismatches 330; Indels 234; Gaps 46;  
Qy 114 RFSTEEDGNGSPENNFLADPSSDLVFTLHNTPTPFQFVSRRSSG-----DILFDTSPSSD 169  
Db 205 RVTRKDKGKIMEN-----NEVPAG-SLGNKCRGLMFVDRLYGTATASVY 248  
Qy 170 SNTYEIFDKQFLQSSALPENRSLNYLIGE-----HTKRSFLI--PGETWTLNAD 219  
Db 249 EN-----YRND-----PDRKEGFYAGVEVCEVDFWSEQNRYKYLILERTGIAMTNTNYD 296  
Qy 220 IGSEN-----PDVNLGSHPFY-----MDVRGSKGNEEAGTTHGVLL-----L 257  
Db 297 NYNTNQSDLIAPGPSDPNFY--IPMYFAAPMVVVGCGSNGSDEQSYSGWFMKNVSYOTYM 354  
Qy 258 NSNGMDVKYEGHRITYNVIG--GVIDL-YVFA-GSPPEMVMMQYTELIG-----302  
Db 355 NTGGSWMNCGEENLAY--MGAQCGPFDOHFVYGDGDLGVQAFSLLOQKEFENQVLNK 412  
Qy 303 RPAPMPYWSFGFHQCRYCY-----KNVSDLEYVVDGYAKAGIPLVEMWTDIDY 350  
Db 413 RAVMPKPYFGYFGVFGIASLLREORPEGGNNIS-VQEI1VEGYGSNNFPLEGLAVDVM 471  
Qy 351 MDGYKDFTLDPVNPREDKMQSFVDTLHKN-----GQKYVL 385  
Db 472 QODLRVFTTKIEFTWANKVGTGGSDNNKSVFEWAHDKGLVCQTNVTCFLRNDNGGADYEV 531  
Qy 386 ---ILDPGIG-----VDSYGTYNRGMEADVFKRNGEPYLGVEWPG-----KYVPPDF 431  
Db 532 NQTLREKGLYTKNDLSLTNTNFTNDG-----PSDAYIGHLDYGGGNCDCALPDM 582  
Qy 432 LNPAATAFW-SNEIKMFOEILPLDGLWIDMMELNSFNITSPSSGSSLLDD-PPYKINSGD 489  
Db 583 GRPGVAEWMGDNYSKLPK--IGLDFVQDMTTPA---MMPHKVGDVADVTRSPYGMNPEND 637  
Qy 490 KRPINK-----TVPATSIHFGNISE---YDAHNLYLLEAKATHQ-----AVVDITG 534  
Db 638 --PSNGRYNMKSYHPQVLVTDMYENHGRPEMFTQRMHAYTLCESTKEGIVANADTLT 695  
Qy 535 K--RPFILSRSTFVSSGKYTAHTGDNAAKWEIDLAYSIPGILNFGLPFIQIPMWGADICGFS 592  
Db 696 KFRSYIISRGYIGNQHFGGMVVDNSSSORYLQMMIANIVNNMNSCLPLVGSIDIGFT 755  
Qy 593 HDTTEELCR-----RWIOLGAFYPPARDH-----SSLGTARQELVLDMSVASSARKVL 640  
Db 756 SYDGRNVCPGDLWVRFVQAGCLLPWRNHYGRVLEGKQEGKYQELMYKDEMATLRKFI 815  
Qy 641 GLRMRLPLHLYTLMYEAHVSGNPIARPL-FFSFQDTKTYEIDSOFLI-----GKSIMVSP 695  
Db 816 EFRYRWQEVLYTAMYQNAAFGKPIKAASMYDNDNRVGAQ--DDHFLGCGHDGYRILCAP 874  
Qy 696 ALKQGAVAVDAYFPA-----GNWFDLFNYSFVAGGDSGKHVR-LDTPADHVNVHVREG 747  
Db 875 VVWENTTSRDLYLPVLTWKYKFGPDYDTKRLDSAL--DGGQMIKNYSVPQSDSPIFVREG 932  
Qy 748 SIVAMQ-----GEALTTRDARKTPYQLLVASRLNISGELFLDDG-----788  
Db 933 ALPTRYTLDSGNSKSMNTY-TDKDPLVEFEVPLGNRAGDGCYLLDDCGITTTDAEDHGKFS 991  
Qy 789 -----ENLRMGAGGGRNDWTLVKFC-----YVTGKSVLVRSEVNVPEYASKMKWSIG 836  
Db 992 VINVEALRQV-----TTTIKFAVDYQYVDFGPFYVR--IRNLTTASKINSSG 1039

## RESULT 9

US-08-633-768A-1  
; Sequence 1, Application US/08633768A  
; Patent No. 6013504  
; GENERAL INFORMATION:  
; APPLICANT: YU, SHUKUN  
; APPLICANT: BOUSEN, KIRSTEN



REFERENCE/DOCKET NUMBER: DY09.001C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 714-760-0404  
TELEFAX: 714-760-9502  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1092 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-275-608-3

Query Match 7.5%; Score 356.5; DB 4; Length 1092;  
Best Local Similarity 22.7%; Pred. No. 2.1e-24;  
Matches 183; Conservative 116; Mismatches 278; Indels 229; Gaps 43;  
Qy 124 SPENNFLADSSDLVFTLHNTTPFGFSVR--RSSGDILF--DTSPDSSDSNTY--FIK 177  
Db 181 SPE-GFLSFETKDL-----NVIIYGNCKMRVTKKDGVLVMEDECNQSQSGNKCRLMYV 234  
Qy 178 DOFLQSLALPENRSN-----LYGIGEHT-----KRSFLIPGQETW 213  
Db 235 DRL--YGNATIASVQTNFHKDTSRNEKFGAGVNCRYEEQGKAPTYYVLERS-----GLAM 287  
Qy 214 TLWNAJ-IGSENPDV--NLGSHPFY-----MDVRGSKGN-----BE 247  
Db 288 TNYNDNLNLYNQPDVPPGPDHNYIIPMYAAPLWVGCGAGTSKQSYSGHFMNVSQ 347  
Qy 248 AGTHGVLLNSGMDVKYEGHRTYNYVIGVIDLYVA-GPSPMYMNOYTELIG---- 302  
Db 348 SYMNTGDTAMCGOENLAYMGAQ-----YGPFDQHFVYGDGDLGVKAFSPLOGKEFE 402  
Qy 303 -----RPAPMPYSGFHGCRYY-----KNVSDLEYVDGYAKAGIPIELVMW 345  
Db 403 DKLNKRSWMPKPYVFGFGVFGALSKLLKONLPAGENNISVQIEVGYQDNYPFEGLA 462  
Qy 346 TDIDYMGYKDFTLDPVNPFPEDKM-----QSFVDTLHNGKQYVLIIDPDTGV---- 393  
Db 463 VDVMQDLDLRFVTTKPEYMSAMVGECDPNRSVFEMAHDRG----LVQTNVTCFLRN 518  
Qy 394 DSSVGYT--NRGM-EADVFIKRN-----GEPYLGVEWPG-----KVFPDP 431  
Db 519 DNSGKPYEVNQLREKQLYTNDLNTDFGTTSDGPGDAYIGHLDYGGVGECDALFPDW 578  
Qy 432 LNPAAATFW-SNEIKMFOEILPLGLTDM-----NELSNFITSPLSSGSLDDPZYK 483  
Db 579 GRPDVAQWGENYKLF--IGLDFVQDQMTVPAMPHRLGDAVKN--NSGSSAPGW-- 632  
Qy 484 INNSGDKRPINK-----TVPATSIHFG-----NISEYDAHNLGLLEAKATHQA 528  
Db 633 --NEND--PSNGRYNWKSYHFQVLVDMRYGAEGREPMVSORNIH-AYTLCESTRREGI 687  
Qy 529 VVDITG-----KRPFILSRSTFVSSGKYTAHTWGTGNAKWEIDLAYSIPGILNFGLGIPMV 584  
Db 688 VGNADSLTKFRSRIIISRGYIGNHQHFGGMVWGNSATESYLOKMLANIINMWSCLPLV 747  
Qy 585 GADICGFSH-----DTTEELCRRWIQLGAFYFPAROH-----SSIGTAROEIYLWDS 631  
Db 748 GSDIGGFTQYNDAGDPTPEDLKYRFVQAGCLLPWRNHYDRWIESKHKGYQELMYPG 807  
Qy 632 VASSARKVLGLRMLRLLPHLYLTMYEAHVSGNPI--ARPLFFSPDQTKTYEI--DSQFJ-I- 687  
Db 808 OKDTLKKEFEFRYRQWELVYLTAMYQNTATTGPIIKAAPM--YKNDVNVYKSNQCHFLUG 864  
Qy 688 ---GKSIWSPALKOGAVAVDAYPPA-GNWF-----DLENYSPAVG 724  
Db 865 GHDGYRILCAPVRENATSRVLYLVYSKFKFGDFDTKPLENEIQGGQTLNYA---- 920  
Qy 725 GDSGKHVRDLTPADHNVHVRREGSIV 750  
Db 921 -----APLNDSPIFVREGTIL 936

RESULT 11  
US-09-275-608-4  
Sequence 4, Application US/09275608  
Patent No. 6541237  
GENERAL INFORMATION:  
APPLICANT: YU, Shukun  
APPLICANT: JSEN, Kirsten  
APPLICANT: MARCUSSEN, Jan  
TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE AND  
TITLE OF INVENTION: ITS USE IN THE PRODUCTION OF 1,5-ANHYDROFRUC  
TITLE OF INVENTION: TOSE  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/275,608  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP95/02172  
FILING DATE: 06-JUN-1995  
APPLICATION NUMBER: 08/836,156  
FILING DATE: 15-APR-1997  
APPLICATION NUMBER: PCT/EP94/03397  
FILING DATE: 15-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Altman, Daniel E  
REGISTRATION NUMBER: 34,115  
REFERENCE/DOCKET NUMBER: DY09.001C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 714-760-0404  
TELEFAX: 714-760-9502  
TELEX:  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 570 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-275-608-4

Query Match 7.0%; Score 333.5; DB 4; Length 570;  
Best Local Similarity 23.8%; Pred. No. 1e-22;  
Matches 140; Conservative 84; Mismatches 193; Indels 171; Gaps 27;  
Qy 213 MTLWNAJ-IGSENPD-----VNLGSHPFYMDVRGSKGNEEAGTTHGVLLN 258  
Db 1 MTNYNDNLNLYNQPDILPPGHSDSDPDYIIPMYFAAPMWI-AHGYRGTSQD-QSYGWFLLN 58  
Qy 259 -----SNGMDVKYEGHRTYNYVIGVIDLYVF-AGPSPMYMNOYTELIG-- 302  
Db 59 VSQSYTNTGDDAWAQKDLAYMGAQ-----CGPFDQHFVYEAGDGLDVDTAFSYLQKE 113  
Qy 303 -----RPAPMPYSGFHGCRYY-----GYKQVSDLEYVDGYAKAGIPIE 342  
Db 114 YENQGLNRSAMPKPYVFGFGVFGATSLLRDNLNLPAGENNVLS-LEEIVGYQNQNVPE 172  
Qy 343 VMWTDIDYMGYKDFTLDPVNPFPEDKMOSFVDTLHNGKQYVLIIDPDTGVDSYGTNR 402  
Db 173 GLAVDVMQDLDLRFVTTKPEYMSAMVGECDPNRSVFEMAHDRG-----AHDR 217

QY 403 GM-----EADVFIKNGEPYLGVE-----WPKVY----- 427  
Db 218 GLVCQNTVTCFLKNEKXP--EVNQSREKQLYTKSDLSLDNIDFGTTDPGSDAIGHLD 275  
QY 428 -----FPDPLNPAATFW-SNEIKMFOEILPLDLGLWIDMNLNFIITSPLSSGSS 476  
Db 276 YGGVECDALFPDGRPDVAQWGDNDYKCLFS--IGLDFVQDM-----TVPAMMPHR 326  
QY 477 LDPPPKINNSG-----DKRPINK-----TVPATSIHFGN-----ISEYDAH 514  
Db 327 LGDPVG--TNSGETAPGWNKDPNSGRYLNKWSYHPQVLVTDMRYDDYGRDPVITQRNLH 384  
QY 515 NLYGLLEAKATHQAVVDITG----KRPFLSRSTFVSSGKYTAHWTGDNAAKWEOLAYS 570  
Db 385 -AYTLCSTEREGIVGNAUSLTKFRSYIISRGYIGNQHFGGMVGDNSSTEDYLMVW 443  
QY 571 PGLNGLFGIPMVGADICGFS-HD-----TTEELCRRTWOLGAFYPPARDH-----SS 618  
Db 444 INVNNMNSGVPLVGSIDIGFTEHDKRNPCPTDLMRFVQAGCLLPWFRNHYDRWIESKK 503  
QY 619 LGTARQELYLWDSVASSARKVLGLRMLLPHLYTLMYEAHVSGNPIAR 666  
Db 504 HGKNYQELMYRDLHALRSFVELRYRMOEVLTYAMYQNALNKPIIK 551

## RESULT 12

US-09-634-238-331  
; Sequence 331, Application US/09634238  
; Patent No. 6544772  
; GENERAL INFORMATION:  
; APPLICANT: Glenn, Matthew  
; APPLICANT: Havukkala, Ilkka J.  
; APPLICANT: Bloksberg, Leonard, N.  
; APPLICANT: Lubbers, Mark W.  
; APPLICANT: Dekker, James  
; APPLICANT: Christensson, Anna C.  
; APPLICANT: Holland, Ross  
; APPLICANT: O'Toole, Paul W.  
; APPLICANT: Reid, Julian R.  
; APPLICANT: Coolbear, Timothy  
; TITLE OF INVENTION: Polynucleotides, materials incorporating  
; FILE OF INVENTION: them and methods for using them.  
; FILE REFERENCE: 11000.104301  
; CURRENT APPLICATION NUMBER: US/09/634,238  
; CURRENT FILING DATE: 2000-08-08  
; NUMBER OF SEQ ID NOS: 422  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 331  
; LENGTH: 390  
; TYPE: PRT  
; ORGANISM: Lactobacillus rhamnosus  
US-09-634-238-331

Query Match 4.0%; Score 190; DB 4; Length 390;  
Best Local Similarity 21.2%; Pred. No. 2.7e-09;  
Matches 102; Conservative 57; Mismatches 191; Indels 130; Gaps 20;  
QY 266 YEGHRITNYIGGVIDLYVPAGSPBWM-----NOYTELIGRPAPMPYWSGPHQC 317  
Db 8 YEHLNAYN-----RDTWV---VPBGTGSAIKFEDGKYKEY--QPGLIPAGQGIGKES 57  
QY 318 RYGVKN--VSDLEVVYVDGAKGIPLEVMWMTDIDYMDGYKDFLDPVNFPEDMQMSFVT 375  
Db 58 LNCGLNNYQFSARAVIDRYQKNDMPGLCWFLPNDGYGAGYQ-----TDT 101  
QY 376 LHKNGQKYVLIIDPGIVSSYTY--NRGMEADVFIKNGEYPYLGEVWPGVKYFPDFLN 433  
Db 102 LAGNLQNL-----KSFADYADQHGVAATGLMTQNLSPV-----DPAN 138  
QY 434 PAA-ATFWSNEIKMFOEILPLDLGLWIDMNLNFIITSPLSSGSSLDPPYKINNSDKRP 492  
Db 139 PKPDRDFAKEVAIGKALKTDAWV-----GSGYS----- 169

QY 493 INNKVTPATSIHFGNISYDAHNLYLEAKATHQAVVDITG--RPFILSRSTFVSSGK 550  
Db 170 -----LDGLAKADA---MMTQVKGDSLRFALITLDCWAGTQR 205  
QY 551 YTAHTGCD-NAAKWEDLAYSIPGILNGLFGIPMVGADICGFSHDTTTEELCRRWIQLGAF 609  
Db 206 YAGVMTGDTQGTQGWEXYIRPHIPYITGTLGSGQPVGSDMDGIFGGGNPIVTRDFQWKAF 265  
QY 610 YPFARDHSSIGTARQELYLWDSVASSARKVLG-LRMRLPLHLYTLMYEAHVSGNPIARPL 668  
Db 266 TPIQLNMDGNGANPKTPFSFDQQTATINRAYNKQKTMLEFNYNTASAQSVDFGKPMVRGL 325  
QY 669 PFSFPQDTKYE--IDSQFLIGKSIWSPALKQGA-----VAVDAYPPAGN--WFDLF 717  
Db 326 FLDYPNPEAYTDLVKXEYLWGNFLVAPIYQNTADEKGNDRVNGIYLPDKQQWIDYY 385

## RESULT 13

US-09-437-054A-10  
; Sequence 10, Application US/09437054A  
; Patent No. 6316698  
; GENERAL INFORMATION:  
; APPLICANT: Allen, Stephen M.  
; APPLICANT: Kinney, Anthony J.  
; TITLE OF INVENTION: Plant Alpha-Glucosidase II Homologs  
; FILE REFERENCE: BB1273 US NA  
; CURRENT APPLICATION NUMBER: US/09/437,054A  
; CURRENT FILING DATE: 2001-05-14  
; PRIOR APPLICATION NUMBER: 60/107,909  
; PRIOR FILING DATE: 1998-No. 6316698member-10  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 10  
; LENGTH: 150  
; TYPE: PRT  
; ORGANISM: Triticum aestivum  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (73)-(74)  
; NAME/KEY: UNSURE  
; LOCATION: (102)  
; NAME/KEY: UNSURE  
; LOCATION: (123)  
; NAME/KEY: UNSURE  
; LOCATION: (141)  
; NAME/KEY: UNSURE  
; LOCATION: (143)  
US-09-437-054A-10

Query Match 3.9%; Score 185; DB 4; Length 150;  
Best Local Similarity 29.9%; Pred. No. 1.4e-09;  
Matches 52; Conservative 23; Mismatches 71; Indels 28; Gaps 5;  
QY 642 LRMELLPHTLYMYEAHVSGNPIARPLFFSPPOSTKTYEIDSOFLIGKSIWSPALKQGA 701  
Db 4 MYSLLPYYSILFQEAASVTGVPVMPPLWLEFPDDKETYNNGEAFMVGPSILAQGIYEEGQ 63  
QY 702 VAVDAYPPAGN-WPDLFNYSFAVCGSDSKHVRDTPADRYNVHVHREGSIVAMQGEALTTR 760  
Db 64 KSVSVYLFQXXLWYDLRNGSPYKGSVSHK---LQVSEDSIPXFORSGTIV----- 110  
QY 761 DAKRTPQLLVAVASRLNLSGELFLDDGELNRMGAGGGRDWTLVKPRCYVTGK 814  
Db 111 -PRDRFRILTOX-----VNDSYTLVIGL---NNSWLXKXTYYDDGK 150

## RESULT 14

US-09-367-895-41  
; Sequence 41, Application US/09367895  
; Patent No. 6483009  
; GENERAL INFORMATION:  
; APPLICANT: POULSEN, PETER





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:19:13 ; Search time 48.6618 Seconds  
(without alignments)  
2942.169 Million cell updates/sec

Title: US-10-043-418-4

Perfect score: 4751

Sequence: 1 MSSLHFPNIFIVVVFSL.....EVSLSLLVGKFFEMRLRLT 902

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq\_19Jun03.\*

```
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	2706	57.0	901	23	AAU97731
2	2706	57.0	901	23	ABO9151
3	2703	56.9	901	23	AAU97732
4	2703	56.9	901	23	ABO9152
5	2451	51.6	877	19	AAW59040
6	2350.5	49.5	682	21	AAV51670
7	1472	31.0	953	23	ABR57174
8	1424	30.0	922	20	AAW88044
9	1423	30.0	985	14	AAU42995

10	1423	30.0	985	14	AAU42214
11	1405.5	29.6	963	20	AAU49895
12	1360	28.6	1827	20	AAW74090
13	1360	28.6	1827	22	AAU09028
14	1360	28.6	1829	22	ABG14994
15	1357	28.6	985	18	AAW15191
16	1347	28.4	958	9	AAU81181
17	1273.5	26.8	1070	13	AAU07575
18	1273.5	26.8	1070	13	AAU1521
19	878.5	18.5	914	23	ABP52437
20	877.5	18.5	912	23	ABG9613
21	876.5	18.4	914	23	ABG9611
22	875.5	18.4	914	23	ABG9612
23	868	18.3	763	22	ABR50119
24	864	18.2	941	22	AAU80202
25	864	18.2	944	22	AAU79218
26	863.5	18.2	912	24	AAU79779
27	858.5	18.1	914	23	AAU66094
28	837.5	17.6	967	23	ABG9610
29	833	17.5	919	18	AAU18580
30	833	17.5	919	23	AAU66099
31	761	16.0	565	21	AAU58849
32	728.5	15.3	924	22	ABG7421
33	728.5	15.3	924	22	ABR69429
34	713.5	15.0	693	24	ABP96604
35	713.5	15.0	712	24	ABP96616
36	713.5	15.0	718	24	ABP96615
37	713.5	15.0	718	24	ABP96625
38	685	14.4	444	21	AAU42826
39	624.5	13.1	834	11	AAU04869
40	601.5	12.7	235	22	AAU00087
41	505.5	10.6	390	23	ABP51380
42	494.5	10.4	1070	16	AAU77083
43	494.5	10.4	1070	16	AAU72712
44	494.5	10.4	1070	20	AAU88256
45	482.5	10.2	331	22	AAU23490

#### ALIGNMENTS

RESULT 1  
AAU97731  
ID AAU97731 standard; Protein; 901 AA.  
XX  
AC AAU97731;  
XX  
DT 23-AUG-2002 (first entry)  
XX  
DE Common buckwheat alpha-glucosidase polypeptide #1.  
XX  
KW Common buckwheat; alpha-glucosidase; enzyme;  
XX  
KW Fagopyrum esculentum Moeench.  
XX  
OS Fagopyrum esculentum.  
XX  
PN JP2002065273-A.  
XX  
PD 05-MAR-2002.  
XX  
PF 31-AUG-2000; 2000JP-0262104.  
XX  
PR 31-AUG-2000; 2000JP-0262104.  
XX  
PA (NISO ) NIPPON SHOKUJIN KAKO KK.  
XX  
DR WPI; 2002-474198/51.  
XX  
DR N-PSDB; ABR86269.  
XX  
PT A recombinant vector containing alpha-glucosidase gene, and a  
XX  
PF transformant useful for preparation of alpha-glucosidase -  
XX  
PS Claim 2; Page 7-8; 18pp; Japanese.

Aspergillus niger  
Endomycetes fibuligie  
Human hsi protein  
Human sucrose isom  
Novel human diagno  
Aspergillus oryzae  
Sequence of glucoc  
Glucosylase encod  
Alpha galactosidas  
Human carbohydrate  
Human NOV11d prote  
Human NOV11b prote  
Listeria monocytog  
Human protein SEQ  
Human protein SEQ  
Carbohydrate-assoc  
Soybean alpha-gluc  
Human NOV11a prote  
Potato alpha-gluc  
S. tuberosum alpha  
Breast and ovarian  
Drosophila melanog  
Drosophila melanog  
S. solfatarius ma  
Plasmid pNOV4839 m  
Plasmid pNOV4831 m  
Mala fusion protei  
Human ORFX ORF2590  
Alpha-1-6-glucanas  
Alpha-glucosidase  
Human MBDT SEQ ID  
M. vulgaris alpha-1  
Alpha-1-4-glucan l  
Morcella vulgaris  
Novel human enzyme



XX		The invention relates to a recombinant vector containing
CC		alpha-glucosidase gene derived from buckwheat. A transformant containing
CC		the recombinant vector can be used for preparing alpha-glucosidase by
CC		culturing the transformant and isolating alpha-glucosidase from the
CC		CC culture. This sequence represents a buckwheat alpha-glucosidase protein.
XX		
SQ	Sequence	901 AA;
	Query Match	57.0%; Score 2706; DB 23; Length 901;
	Best Local Similarity	58.2%; Pred. No. 2.5e-220;
	Matches 527; Conservative	120; Mismatches 222; Indels 36; Gaps 11;
QY	7 FPNIFVVVFFSLRRSQVLVEEESTVGXGVYVRVSVDGNSRQVLTAKLDLTKPSVY	66
DB	22 FCSLFVY-----SSSGEVGGYGVRVRAKVADSSNTTATFAKLINASSLY	66
QY	67 APIKISLNHLVSLSTSRRLRIITDSSQORWEIPETVIPRAGN--HSPRRFSTEEDCGN	123
DB	67 GDQIPNLITFTAFEKDYRLRIIRIDAEKPWEIENEVLHRDGSSHGHQP---LDSPTT	122
QY	124 SPENNFIADPSSDLVFLTHNTTPFGFSVRRSSGDIILFDT-SPDSSPSNTYFIFKDQFLQ	182
DB	123 PPSAAVLTHENSOLIFRLHDTPFGFSVTRRSWDLVDFTRSADPETDPVGLVFVKDQVIQ	182
QY	183 LSSNALPENRNLCIGIEHTKRSPFLIIPEGETMTLWNADIGSENPDVNLYGSHPFVMWRGS	242
DB	183 LSSSLPGRRALHYGIGEHKTPFERLAHNQTLLTNADIASYNVDNLYGSEHPFYLDVRA-	241
QY	243 KGNBEAAGTHGVLLNSNGMDVKYEGRHRTYNNVIGGVIDLXFVAGPGEPMVMNOYTELIG	302
DB	242 ----PLGTSGVLLNSNGMDVEYTGDRITYKVIGGIIDLXVFAGPTDEVVQQYTELIG	297
QY	303 RPAMPYWSFGFHOCRYGYKNVSOLEYVVDGYAYAGIPLEVMYTDDIDYMGCKDTLDLPV	362
DB	298 RPAMPYWSFGFHOCRYGYRVNVSVVENVKAYATMRIPLEAIWTDDIDYMEANKDFTVPDV	357
QY	363 NFPEDKMOSFVDTLHKNCOKVVLILDPCIGYD-VSYGTXYARGMEADVPIKENGEPYLGEV	421
DB	358 NFPUDKMQRFVNLKHNGCQKVAILDPCGINNTTYGTGFQAMKADIFIKRGQGPYQGEV	417
QY	422 WPKVYVPPDFLNPAATAFWSEIKMFQBIPLDGLJWDNMNELSNFITSPSSSGSSLDDPP	481
DB	418 WPGVYVPPDFLNPKTTTFWSEIQTFENALPVCGJLMDMEVSNFISGPPIPCSELDNP	477
QY	482 YKINNSDKGRPINNKTVPATSIHFGNISEYDAHNLYLEJEAKATHQAVDITGKRPFILS	541
DB	478 YVINNSGRRPINKEKTIPVSSVHHYCNVSDYVNHVNLGYLEAIAATNALVKVTKQRPFVLS	537
QY	542 RSTFVSSGKYTAHTWGNAAKWEDLAYSIPIGLNFGLFGIPMWGADICGFSHDTEELCR	601
DB	538 RSTFIGSGKTAHTWGNAATWTNMAASIPILDFGLFGIPMIADICGFAYTTTEELUCR	597
QY	602 RWIQLGAFYPFARDHSLGTARQELYLWDSVASSARKVLGERMLLPHLYTLMYEAHVSG	661
DB	598 RWIQLGAFYPPSRDHSDNLTEPQELTWGVSVTETARKVLGRYLRLLPYLYTMLYEAHKKG	657
QY	662 NPIARPLFFSFQDTKYIEDSOFLGKSIMVSPALQCAVAVDAYPPAGHWFLFNYSF	721
DB	658 TPIARPLFFSFPPDNPTVGIDAQFLVGLGVNVSPVLTCGATSVTAYFPFSGMWFNLFDTK	717
QY	722 AVGG-DSGKHVRLDTPADHVNHVHREGSIVAMOQEALTTRDEARTPYQLLVVASRLENIS	780
DB	718 TVSSPTNGSFVTLDAPLEEINVHVREGSILLAQEMTTREARNTPFELVVVISDSNGGS	777
QY	781 --GELFLDDGENLRMGAGGGRNDWLKFRCYVTGKS-VVLRSEVNPDEYASKMKWSIGK	837
DB	778 SIGSVFLDNGVDIEWGDGGR--NSLVTFSSAGLIIGNKKVTTTSSVVGGRFALSQGWKISK	835
QY	838 VTFVGFENVENKTYEVRTSERLRASPRLISLIKTVSDNDDDRFLSVESKSLILVKKFKPEM	897
DB	836 VTIIGLRSGSVKGYTMLSIGRVVTSYKVGARKGS-RGTGKFDDVVEIPLNLSLLVGRNFLK	894

QY 303 RPAPMYWFGFHQCRYGKYNVSDLEYVVDGYAKAGIPLEVMMWTDIDYMDGKDFLDPV 362  
DB 298 RPAPMYWFGFHQCRYGKYNVSVVENVKAYATMRIPLEAIWTDIDYMEANKDFVDPV 357  
QY 363 NPPEDKMQSFVDTLHKNQKQYVILDPGIGVD--SSYGTYNRGMADVFIKRNGEPPYLGCV 421  
DB 358 NFPDLDKMQRFVNLKHNQKQYVAILDPGININTTYYGTQRAKADIFIKRQCEPYQGEV 417  
QY 422 WPGKVVPDFLNPAAATFNSNEIKMFOEILPLDGLWIDMNELSNFTTSLSSGSSLDLDP 481  
DB 418 WPGVTFPDPFNPKTTFIWEISEIOTFNALPVDGLWIDMNEVSNFISPPIDSLDNP 477  
QY 482 YKINSGDKRPINNKTPATSIHFNGISEYDAHNLYGILLEAKATHQAVVDITGKRPFILS 541  
DB 478 YVINNSGRRPINEKTI PVSSVHYGNVDYVNVHNLGYLEAIATNVALKVKTKQRPFVLS 537  
QY 542 RSTFVSGKYTAHWTGDNAAKWEDLAYSPGILNPGIFGIPMGADICGFSHDTTEELCR 601  
DB 538 RSTFIGSKYTAHWTGDNAAATWNNAASTPILDFGLGIPMGADICGFAITTEELCR 597  
QY 602 RWIQLGAFYFPARDHSSLSGTARQELYLWDSVASSARKVLGLRMLLPHLYTLMYEAHVSG 661  
DB 598 RWIQLGAFYFPARDHSDNLTEPQELTQWGSVTETARKVLGLRYLLPYTYTLMYEAHKG 657  
QY 662 NPIARELFESFPQOTKTYEIDSOFLIGKSIWSPALKQGA VADAYFPAGNMFDLNFYSF 721  
DB 658 TPIARELFESFPNDPNTYGIDAQFLVGKGVWSPVLTCGATSVTAYFPGSNWFLFDYTK 717  
QY 722 AVCG--DSGKHVRLDTPADHNVHVRGSIWAMQGEALTTTRDARKTPYQLLVASRLNIS 780  
DB 718 TVSSPTNGSVTLTDAFLBEEINHVHVRGSIWALQGEAMTTREARNTPELVVWISDNGS 777  
QY 781 --GELFLDDGENLRMGAGGGRDWTLVKFCYVTGKS--VVLRSVNVPEYASQKWSICK 837  
DB 778 SIGSVFLDNGVDIEMGDGGR--WSLVTFSAGLIGNNKVTITSSVNGRFAISQGHKISK 835  
QY 838 VTFVGFENVENKTYEVRTSERLSRSPRISLIKTVSDNDPDRFLSVSVSKSLSLVAGKKFEM 897  
DB 836 VTILGLSRGQVKGTYMWSIGRVVTTISVKGARKCS--RGTGKFDVVEIPNLSLLVGNRFXL 894  
QY 898 RLRLT 902  
DB 895 DIQIT 899

RESULT 3  
AAU97732  
XX  
AC AAU97732 standard; Protein; 901 AA.  
XX  
DT 23-AUG-2002 (first entry)  
XX  
DE Common buckwheat alpha-glucosidase polypeptide #2.  
XX  
DE Common buckwheat; alpha-glucosidase; enzyme;  
KW Fagopyrum esculentum Moeench.  
KW  
XX Fagopyrum esculentum.  
OS  
XX JP2002065273-A.  
XX  
XX 05-MAR-2002.  
XX  
XX 31-AUG-2000; 2000JP-0262104.  
XX  
XX 31-AUG-2000; 2000JP-0262104.  
XX  
XX (NISO ) NIPPON SHOKUIN KAKO KK.  
XX  
XX WPI; 2002-474198/51.  
DR N-PSDB; ABX86270.

XX A recombinant vector containing alpha-glucosidase gene, and a  
PT transformant useful for preparation of alpha-glucosidase -  
XX  
XX Claim 2; Page 8-9; 18pp; Japanese.  
XX The invention relates to a recombinant vector containing  
CC alpha-glucosidase gene derived from buckwheat. A transformant containing  
CC the recombinant vector can be used for preparing alpha-glucosidase by  
CC culturing the transformant and isolating alpha-glucosidase from the  
CC culture. This sequence represents a buckwheat alpha-glucosidase protein.  
XX  
SQ Sequence 901 AA;  
Query Match 56.9%; Score 2703; DB 23; Length 901;  
Best Local Similarity 58.6%; Pred. No. 4.5e-220;  
Matches 529; Conservative 119; Mismatches 228; Indels 26; Gaps 12;  
QY 10 IFIVVVVFFSRSSQVLEBEESTVVGYYVVRVGVDSNRQVLTAKLDLIKPSVYAPD 69  
DB 15 LLAATLLFCSL--FVWSESE--VVGYYRVVRKAVDSSSNTLTAFKLINASSLYQD 69  
QY 70 IKSINLHVLSLTSERLIRITDSSQORWEIPETVPRAGN---HSPRRPSTEEDCGNSPE 126  
DB 70 IPNLTFTATPEKDYRLRIRITDAEKPRWEIPNVLHRDGSQSHGQP---LDSRPTTPPS 125  
QY 127 NNFLADPSSDLVFTLHNTTTPFGSVSRSSGDIILFDT--SPDSSDSNTYFIFKQDFLOLSS 185  
DB 126 AAVLTHPNSDLIFRLHDTNPFGSVTRRSTNDVLFDTRSADPETDPVCLVKDQYIOLSS 185  
QY 186 ALPNRNLGYIGGHTKRSFRLIPGEMTMTLNADIGSENPVNLYGSHPPFMDVRGSKGN 245  
DB 186 SLPADRSNLGYIGGHTKPTFLARNQTLTLNADIASYNDNLNLYGSHPPFYLDVRA---- 241  
QY 246 BEAGTHCVLLNNGMDVYEGHRTYVNVIGVLDLYVVFAGSPSPVMQNYCYTELIGRPA 305  
DB 242 -PLGTSNGLLNNGMDVEYTGDKITYYKVGIVDLYVFEPTDVEVYQVYTELIGRPA 300  
QY 306 PMPYMSFGFHQCRYGKYNVSDLEYVVDGYAKAGIPLEVMMWTDIDYMDGKDFLDPVNF 365  
DB 301 PMPYMSFGFHQCRYEYRNISVVENVKAYSITMRIPLEAMWTDIDYMEANKDFTVDPVNF 360  
QY 366 EDKMQSFVDTLHKNQKQYVILDPGIGVD--SSYGTYNRGMADVFIKRNGEPPYLGCVW 424  
DB 361 LDKQRFVNLKHNQKQYVAILDPGININTTYYGTQRAKADIFIKRQCEPYQGEVW 420  
QY 425 KVFYFDFLNPAAATFNSNEIKMFOEILPLDGLWIDMNELSNFTTSLSSGSSLDLDPVYKI 484  
DB 421 PVYFDFLNPKTTFIWEISEIOTFNALPVDGLWIDMNEVSNFISPPIDSLDNP 480  
QY 485 NNSGDKRPINNKTPATSIHFNGISEYDAHNLYGILLEAKATHQAVVDITGKRPFILSRST 544  
DB 481 NNSGRRPINEKTI PVSSVHYGNVDYVNVHNLGYLEAIATNVALKVKTKQRPFVLSRST 540  
QY 545 FVSSKQYTAHWTGDNAAKWEDLAYSPGILNPGIFGIPMGADICGFSHDTTEELCRRI 604  
DB 541 FIGSKYTAHWTGDNAAATWNNAASTPILDFGLGIPMGADICGFAITTEELCRRI 600  
QY 605 QLGAFYFPARDHSSLSGTARQELYLWDSVASSARKVLGLRMLLPHLYTLMYEAHVSGNPI 664  
DB 601 QLGAFYFPARDHSSLSGTARQELYLWDSVASSARKVLGLRMLLPHLYTLMYEAHVSGNPI 660  
QY 665 ARPLFFSFPPQTKTYEIDSOFLIGKSIWSPALKQGA VADAYFPAGNMFDLNFYSFV 724  
DB 661 ARPLFFSFPPNDPNTYGIDAQFLVGKGVWSPVLTCGATSVTAYFPGSNWFLNFYTKVS 720  
QY 725 GD--SGKHVRLDTPADHNVHVRGSIWAMQGEALTTTRDARKTPYQLLVASRLNIS--G 781  
DB 721 SPANGSVFTLDAPLKEINHVHVRGSIWALQGEAMTTREARNTPELVVWISDNGSSIG 780  
QY 782 ELFLDDGENLRMGAGGGRDWTLVKFCYVTGKS--VVLRSVNVPEYASQKWSICKTFF 840  
DB 781 SVFLDNGVDIEMGDGGR--WSLVTFSAGLIGNNKVTITSSVNGRFAISQGHKISKVTI 838





Query Match		49.5%; Score 2350.5; DB 21; Length 682;
Best Local Similarity		63.5%; Pred. No. 2.5e-190;
Matches 432; Conservative 104; Mismatches 133; Indels 11; Gaps 3;		
QY	225 PDVNLGSHPFYMD-----VRSGKNEEAGTTHGVLLNSNGMDVKYEGHRTITVYIGGV	279
DB	5 PRVHPSQHPIQLHRPAPALHGRGYSFRYPAGVSHGVLLNSNGMDVYVYGRISYKVIIGGL	64
QY	280 IDLYVFAGSPSEVMNQYTELIGRPAMPYMSFGHOCRYGKXVSDLEYVVDGYAKAGI	339
DB	65 IDLYFFAGSPSEVMVQVDTQLIGRPAAMPYMSFGHOCRWGKXNDVDELVDSDYAKSRI	124
QY	340 PLEVWMTDIDYMOCKDFTLDPVNFPEBKQSFVDTLHKQCKYVLLDPGIGVDSYGT	399
DB	125 PLEVWMTDIDYMOCKDFTLDPVNFPEBKQSFVDTLHKQCKYVLLDPGIGVDSYGT	184
QY	400 YNRGMEADVIFKRNGEYPVLGEVWPKYVFPDFLNPAAATFWSNEIKMFQELPLDGLWID	459
DB	185 YRRGMEADVIFKRNMPYQGVWPGVNYYPDFLNPATEVFWNEIEKFQDLVFPDGLWD	244
QY	460 MNELSNFITSPLSGSSLDPPPYKINNSGDKRPINNKTVPATSIHFNGISEYDAHNLYGL	519
DB	245 MNELSNFITSPTSPSTFDDPPPYKINNSGDHLPINRYTVPATSTHFGDTMEYVHNLYGL	304
QY	520 LEAKATHQAVVDITGKRPFILSRSTFVSSGKYTAHMTGDNAAKWEDLAYSIPIGLNGLF	579
DB	305 LESRATYSALVNTGKRPFILSRSTFVSSGKYTAHMTGDNAAKWEDLAYSIPIGLNGLF	364
QY	580 GIPVAGADICGFSHDTTEELCRRIQLGAFYPPFARDHSSLGTAQELYLMDSVASSARKV	639
DB	365 GIPVAGADICGFSNTTEELCRRIQLGAFYPPFARDHSSAKDITPQELYSNDSVAARAKV	424
QY	640 LGLRMRLPHLYTMYEAHVSNCNPIARPLFPSPQDTKYTEIDSQFLIGKSIWVSPALKQ	699
DB	425 LGLRYQLLPYLMYEAHNGTPIARPLFPSPQDTKYTEIDSQFLIGKSIWVSPALKQ	484
QY	700 GAVAVDAYFPAGNWFDFLNTYSFVAGDSDGKHVRLDTPADHVNHVHVSIVAMQGEALT	759
DB	485 GATSVDAYFPAGNWFDFLNTYSFVAGDSDGKHVRLDTPADHVNHVHVSIVAMQGEALT	544
QY	760 RDARCTPYQLLVASRLNISGELFLDDGGENLRMGAGGGRNDWTLVKFRCTVYTKSVYLR	819
DB	545 QAAQRTAFKLLVLLSSKNSKSTGELFVDDDDDEVQMGREGGR--WTLVKFNISIIIGNKIVK	602
QY	820 SEVNVPEYAKKWSIGKTYFVGFEENVENKTYEVTSERLSPRSLIKTVSGNDPPRF	879
DB	603 SEVNVGRYALDQGLVLEKVTLLGFENYRGLSKSYELVGSQQGNTMTK----ESLKQSGQF	658
QY	880 LSVEVSKLSLVLGKKFEMRL 899	
DB	659 VTMEISGMSILIGKEFKLEL 678	
RESULT 7		
ABB57174		
ID	ABB57174 standard; Protein; 953 AA.	
XX		
AC	ABB57174;	
XX		
DT	07-MAR-2002 (first entry)	
DE	Mouse ischaemic condition related protein sequence SEQ ID NO:422.	
XX		
KW	Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;	
XX	vasospastic ischaemia; ischaemic condition; ischaemic disease.	
OS	Mus musculus.	
PN	W0200188-88-A2.	
XX		
PD	22-NOV-2001.	
XX		

PF	18-MAY-2001; 2001WO-JP04192.	
XX		
PR	18-MAY-2000; 2000JP-0145977.	
XX		
PA	(UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.	
XX		
PI	Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;	
XX		
DR	WPI; 2002-034733/04.	
DR	N-PSDB; ABI99468.	
XX		
PT	Examining the ischemic condition (e.g. occlusive ischemia) by measuring	
PT	expression levels of particular genes defined in the specification or	
PT	by determining the expression profile of a gene group comprising these	
XX	genes -	
PS	Claim 2; Page 1154-1158; 2690pp; English.	
XX		
CC	The present invention describes a method for examining ischaemic	
CC	conditions, comprising measuring the expression levels of particular	
CC	genes (1) in a test sample or determining the expression profile of a	
CC	gene group in the sample comprising genes selected from (1). The method	
CC	is useful for examining the ischaemic condition (e.g. compressive	
CC	ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring	
CC	expression levels of particular genes (ABI99202 to ABI99912, encoding	
CC	the protein sequences in ABB57020 to ABB57374) or by determining the	
CC	expression profile of a gene group comprising these genes. The	
CC	expression levels or expression profiles produced by these genes are	
CC	used as an indicator when screening for ischaemic condition-improving	
CC	drugs or therapeutics for ischaemic diseases. ABI99913 and ABI99914	
CC	represent PCR primers for a mouse ischaemic condition related sequence,	
CC	which are used in the exemplification of the present invention.	
XX		
SQ	Sequence 953 AA;	
Query Match		31.0%; Score 1472; DB 23; Length 953;
Best Local Similarity		37.4%; Pred. No. 1.8e-115;
Matches 331; Conservative 139; Mismatches 314; Indels 102; Gaps 25;		
QY	42 RSVGVDSNRQVLTAKLDLKPSSVYAPDKSLNHVSLETSELRIRITDSSQQWEIP- 100	
DB	137 RLENLSSTESGTYATLTRTSP--TFPPKDVLTQLQELVLMETDSRLFKIKDPASKRYEVL 195	
QY	101 EVVIRAGNHSPPRSTEEEDGNSPENNFLADPSSDLVFTLHNTTTFGFSVSRSSGDI 160	
DB	196 ET--PRVLSQAP-----SPLYSVFSEEPFGVIVRKLGGRL 231	
QY	161 FDTSPDSSDSNTYFIFKQDFQLQSSALPENRSLYGIHEHTKRSPLIPGETMTLWNADI 220	
DB	232 LNTTV-----APLFFADQFLQSTSLP--AQHITGLGEHLSPLMLSTDWARITLNRDT 283	
QY	221 GSENPVNLYGSHPPYMDVRGSKGNEEAGTTHGVLLNSNGMDVKYE--GHRITYNVIGV 279	
DB	284 -PPSQGTNLYGSHPPYAL-----EDGGLAHGVFLNLSNAMDVILQPSPALTWESTOGI 336	
QY	280 IDLYVFAGSPSEVMNQYTELIGRPAMPYMSFGHOCRYGKXVSDLEYVVDGYAKAGI 339	
DB	337 LDVYVFLGPEKSVVQVYLDVVGYFPMPPYWGFLGFLCRWGSSTAIVRQVVENWTRTHF 396	
QY	340 PLEVWMTDIDYMDGYKDFTLDPVNFPEBKQSFVD---TLHKNGQKYVLLDPGI---GV 393	
DB	397 PLDVQWNLVDYMDARRDPT-----FNQDSFADFPDMVRDVHQGRRYWMVDPALSSAGP 451	
QY	394 DSSYGTNRGMEADYFI-KRNGEYLVGEVWPKYVFPDFLNPAAATFWSNEIKMFQELP 452	
DB	452 AGSYRYPDEGLRGVFTNETQPLIGKVCPTTAFPDFTNPETLDMQDMVSEFHXQVP 511	
QY	453 LDGLMTDMNLSNFIITSPSS--GSSLDPPYKINNSGDKRPINNKTVPATSIHFNGISE 510	
DB	512 FDGMMLDMNPSFNRVSGQCGCPNNELENPPYXPGVVG--ILQAATCASHQFLS-TH 568	
QY	511 YDAHNLYGLLEAKATHQAVVDITGKRPFILSRSTFVSSGKYTAHMTGDNAAKWEDLAYSI 570	

Db 569 YNLNLYGLTEAIASSALVKTGTRPFFVLSRSTFSGHGAYAGHTWEDVTSWEHLAYSV 628  
QY 571 PGINLFLGPIPMVGADICGFSHDTTEELCRRIQLGAFYPPFARDHSSLGRTARQELYLWD 630  
Db 629 PDILQNLGLVPLVGADICGFIGDTSEELCVRTQLGAFYPPFMRNENLNSVPQPYRES 688  
QY 631 SVASSA-RKVLGLRMLLPHLYTLMEAHVSGNPIARPLFFSPQDTKTYEIDSQELICK 689  
Db 689 ETAQQAARAFALRYALLPLYLTPHRAHVRGDTVARPLFLEPREDPSTWSVDQLLWGP 748  
QY 690 SIMVSPALQOGAVADVAPAGNWFEL-----FNYSFAVGDSGSKHYR 732  
Db 749 ALBITPVEPKTEVTGYFPKGTWYNNQVSVDSGLTLPSPSSASSAFSAV-OSKQWLT 807  
QY 733 LQTPADHVNHVHREGSIVAMQOGALATRDARKTPYQLLVVASRLENISGELFLDDGENLR 792  
Db 808 LEAPLDTINVHLREGVILPLQGPSLTITTESRKQPMALAVALTASGEADGELFWDDGESLA 867  
QY 793 MGAGGGRDWTLYKFCYVTGKSVLRSEVNPDEYASKWISGKVTTFVGPNV-ENVK 851  
Db 868 VLEPHGA---YTLVTFSS--AKNNTIVNKLVRVTKEGA---ELQLKEVTVLGVATAPTQVLS 919  
QY 852 YEVRTSERLSRISLIKTVDNDPREFLSVEYSKLSLLVGKKFEM 897  
Db 920 NGIPVSNFTYSP-----DNKSUAIPV---SLLMGELFQI 950  
RESULT 8  
AAW88044  
ID AAW88044 standard; Protein; 922 AA.  
XX AC AAW88044;  
XX DT 09-APR-1999 (first entry)  
XX DE An enzyme with sugar transferase activity.  
XX KE Acronium sp. S4G13; glucose transfer; sugar transferase  
XX KW sugar receptor; starch.  
XX OS Acronium sp.  
XX PN JP11009276-A.  
XX PD 19-JAN-1999.  
XX PF 19-JUN-1997; 97JP-0163110.  
XX PR 19-JUN-1997; 97JP-0163110.  
XX PS (KIRI ) KIRIN BREWERY KK.  
XX PA WPI; 1999-145893/13.  
XX DR N-PSDB; AAX04639.  
XX DX  
PT New sugar transferase gene and enzyme - useful for catalysing the  
PT transfer of an alpha-1 right arrow 3 bond to a sugar receptor for  
PT saccharide preparation  
XX PS Claim 1; Pages 13-17; 20pp; Japanese.  
XX SQ Sequence 922 AA;  
CC The present sequence represents a sugar transferase protein of Acronium  
CC sp. S4G13. The protein preferably catalyses the glucose transfer of an  
CC alpha-1 right arrow 3 bond or the glucose transfer of an alpha-1 right  
CC arrow 3 and an alpha-1 right arrow 4 bond to a sugar receptor by  
CC reacting with a substrate selected from starch and its decomposition  
CC products.  
XX  
SQ Query Match 30.0%; Score 1424; DB 20; Length 922;  
Best Local Similarity 35.9%; Pred. No. 2.1e-111;  
Matches 317; Conservative 129; Mismatches 285; Indels 152; Gaps 22;

QY 38 GYVRSVGVDNRQVLTAKLDLI-KPSSVYAPDIKSLNHLVSLTSELRIRITDSSOOR 96  
Db 36 GYAASNVQSDTG--LTADTLAGEPCDAYGEDLKOLILEVYETENRHLVKIQDKGNQV 93  
QY 97 WEIPEVTPRAGNHSPPRFSTEEDGSGSPENNFIADPSSDLVFTLHNTTPFGFVSVRSS 156  
Db 94 YQIPESVFPREGS-----IDPSSSIRFAYAEPEFSFNTIADT 133  
QY 157 GDILFDTSDDSDSNTYFIKDFQLQSSALPENRSLNYGIGEHTKRSFRLIPGETWTLM 216  
Db 134 DEVLFDTSAAAS-----IVFESQYLRITSTIPT-PYLYGLGAHNDPMRLESVGYITFW 186  
QY 217 NADIGSENPVNLXGSHPFYMDVRGSKGNEEAGTTHGVLLNSNGMDVKYR-----GHRI 271  
Db 187 NQDSYGVPGNANLYGSHPVYIDHR-----ETG-THGVLFNSNGMDVLIDEDEEGKYL 239  
QY 272 TYNVIGGVIDLYVAGSPSPENMMQYTELICRPAPMPYWSFGHCRCRYGYNKVDLEYVV 331  
Db 240 EYNTLGGVLDYFFYVGDSPSKAVEYGEIAGRPPMPQYMGFLGFHCCKYGYQDAFMVAEV 299  
QY 332 DGYAKAGIPLBVMWTDIDYMDGYKDFTLDPVNFEDKMQSPVDTLHKNGQKYLILDPGI 391  
Db 300 YNYSQAEIPLBVMWTDIDYMDRRVFTVDPDRFPLKIRAVVDYVZHEHQRYIWMVDPAI 359  
QY 392 GVDSYGTYNRMEADVPIKR-NGEPYLGEYMPGVFPDFLNPAAATFWSNEIKMP--- 447  
Db 360 AYVES-GTLDRGLDDVFLRSNGSVMLGVVMPGVTFPDMFAENITQYWNNEFALPFA 418  
QY 448 QEILPLDGLWIDMNELSNF-----ITSP--- 470  
Db 419 DEGVDDIDGLWIDMNEPSNFCNFCNDNPYEAAGFPPTPPVPVPREPRLPGFACVLPQEG 478  
QY 471 -----LSSGSSLD-----PPY 482  
Db 479 TECEDGETAGSKRDSQPGQLVTRQGFSPRPHRRQYEBGDKGLPGRDOLLPEY 538  
QY 483 KINN-----SGDKRPINKTVPATSIHFGNISYDAHNLGLLEAKATHQAV-VDIT 533  
Db 539 AIHNKAAPRDDWADKGGISNKTNTNVIHQGLAEYDVHNLGAMWSSASRDAMEARRP 598  
QY 534 GKRPILSRSTFVSSGKYTAHTWGDNAAKWEDLAYSIPGLNF-GLFGIPMGADICGFS 592  
Db 599 GLRPFIIITRSTPHAGSKVGLWGLDNLNWNQYRESIRTMLAYTSIFQGMVSDVCVGF 658  
QY 593 HDTTEELCRRIQLGAFYPPFARDHSSLGRTARQELYLWDSVASSARKVLGLRMLLPHLYT 652  
Db 659 GDTNEELCARWASLOAFQTFRNHAQYEAQVQEFQWESVAESARRAIGARYRLDDNYT 718  
QY 653 LMVEAHVSGNPIARPLFFSPQDTKTYEIDSQFLIGKSIMVSPALQOGAVADYFPAGN 712  
Db 719 ALWKQSEGTTPAVVPMFVFPEDKGTLELENOYFVPGVLVAPVVEQGSTVDVYVPEGK 778  
QY 713 -WFDLFNYSAVGDSGSKHVLDTDPADHVNHVHREGSIVAM-QGEALTRDARTPYQLL 770  
Db 779 VFYDWTHE-AIQEGGSGSVTYNTTMIPLFIRGGVILPLRENSAMTTTLRKEKFE 837  
QY 771 VVASRLENISGELFLDDGENLRMGAGGGRDWTLYKFCYVTG 813  
Db 838 VALNDGKAKGELYDDGESLE-----QESYTAVKFE-YAHG 873  
RESULT 9  
AA42995  
ID AAR42995 standard; Protein; 985 AA.  
XX AC AAR42995;  
XX DT 16-MAY-1994 (first entry)  
XX DE Glycosyltransferase.  
XX KW Glycosyltransferase; GT; Ctase; Aspergillus niger; plasmid;



Db 58 DAQSVCPGY----KASKVQHSRGFTASLQAGRPCNVYGTDESUTLSVEYQSDRLNI 113  
Qy 88 RI-----TDSSQQRWE-IPETVIPRAGNHSRRSTEDCGNSPENNFADPS--SDLVFTL 141  
Db 114 QILPETHVSTNAGWYFUSENLVPR-----PRASLNASVSQSDLVFSW 155  
Qy 142 HNTTFFGFSYRSGSDILFDTSPDSSDSNTYFIKQDFLQLSALPENRSLNYGIGEH 201  
Db 156 SNEPSFNKVRKATGALFST-----EGTVLVYENQPIEFVFTALPE--EYNLYGLGEHI 208  
Qy 202 XRSFRLIPGETMTLWNADIGSEPNVNLYGSHPPYMDVRGSKNEERAGT----- 250  
Db 209 TO-FRLQARNALTIYPSDDGTP-IDQNLYGQHPPYLDTRYKGDQRNGSYIPYKSSEADA 266  
Qy 251 -----THGVLLNSNGMDVKEYGHRITYNVIGGVIDLVEFAGSPPEMVMNOY-TELIG 302  
Db 267 SQDYISLSHGVLNSHGLILRLRSQKLIWRTGGGIDLTFYSGPAPADVTROYLSTVG 326  
Qy 303 RPAPMPYMSFGHQCRYGYKXVSDLEYVDGYAKAGIPEVMWTDIDYMDGYKDFTLDPV 362  
Db 327 LPAMQOYNTLGFHQCRGYNWSDLDADVANFEKFEIPELVEYINTDIDYMHGYRNFNDQH 386  
Qy 363 NFPEDKMQSFVTLHKQKXVILDPGI-----GVDSSYGTYNRCMEADVFIKR-NGE 415  
Db 387 RFSYSEGDEFSLKXHSRGYYVPIVDAALYIPNPENASDAYATYDRGAADVFLKQPDGS 446  
Qy 416 PYLGEVWPQKVPYDFDLPNPAATFWSNEIKMFOEILPLDGLWIDMELSNFI----- 467  
Db 447 LYIGAWVPGYTVFPDHPHKAQVDFWANELVWIKKVAFGVWYDMSEVSVFCVSGCGTN 506  
Qy 468 ----- 467  
Db 507 LTLNPAHPSFLPCEPGDIIYDYPEAFNITNATEAASASAGASSQAAATATTTSTVSYL 566  
Qy 468 -TSPSSGSSLDPPYKINNSGDKRPINNKTVPATSIHFGNISEYDAHNLGLLEKATH 526  
Db 567 RTPTPGVRNVEHPYVYVNHQEGHDLSVHAVSPNATHVDGVBEYDVHGLYHQGNATY 626  
Qy 527 QAVVDITG--KRPFILSRSTFVSSGKYTAHTWGNAKWEDLAYSPGILNFGLFQIPMV 594  
Db 527 QGLLEWHSKRRPPIIGRSTFAGSGKWAGHGDNYSKMWYSISQALSFLGIPMF 686  
Qy 585 GADICGSHDTTELCCRWIQLGAFYPFARDHSLGTARQELYLWDSVASSARKVLGLRM 644  
Db 687 GADTCGNGNSDEELCNKWLQSAFFPYRNHNLSTIPQEPYRWASVIEATKSAMRIY 746  
Qy 645 RLPLHLYLMEYHVSQNPAPLFFSPQDTKYEIDSOFLICKSTMVSFALKQGAVAL 704  
Db 747 AILPYFYTLFDLAHTTGSTVMRALSNEFFNDPTLAAVETQFMVGPALMVVPELPLVNTV 806  
Qy 705 DAYFP-AGN--WFDLFNYSFVAGGDSGKRVRLDTPADHVNHVHREGSIVAMOGEALTTR 760  
Db 807 KGVPFGVGHVEVMDWTQA-AVDKPGVNTTISAPLGHIPVYVRGNGNILPMQEPALTTR 865  
Qy 761 DAKTPQOLLVVASRLNENISGEFLDDGENL 791  
Db 866 EAROTPEWALLAALGNSGTASQQLYLDGSGE1 896

RESULT 11  
AA49895  
ID AAY49895 standard; protein; 963 AA.  
AC AAY49895;  
XX  
XX 28-JAN-2000 (first entry)  
Dt  
Df  
DE  
XX Endomyces fibuliger glucoamylase GLA2 protein.  
KW Endomyces fibuliger; glucoamylase; GLA2; hydrolysis; carbohydrate;  
KW Glucose; alcohol; starch; leguminous plant; cooked; oleaginous plant;  
KW roasted; cereal; soya; wheat; rice.

XX Saccharomycopsis fibuligera.  
OS EP959130-A1.  
XX 24-NOV-1999.  
XX 06-MAY-1998; 98EP-0201473.  
XX 06-MAY-1998; 98EP-0201473.  
XX (NEST ) SOC PROD NESTLE SA.  
PA Pridmore RD, Kochhar S;  
PI WPI: 1999-622099/54.  
XX N-PSDB; AAZ32388.  
XX  
PT New recombinant glucoamylase from Endomyces fibuliger, useful for  
PT hydrolyzing carbohydrate-containing materials -  
XX Claim 1; Page 13-15; 20pp; English.  
XX  
CC The present sequence represents Endomyces fibuliger (also called  
CC Saccharomycopsis fibuligera) glucoamylase GLA2. The invention describes  
CC the recombinant GLA2 enzyme, which is useful for hydrolysing  
CC carbohydrate-containing materials, especially materials containing rice  
CC carbohydrates. The glucoamylase is used to produce glucose and alcohol  
CC from starch containing raw materials. GLA2 is useful for hydrolysing  
CC carbohydrate containing materials e.g. mixtures of a source of proteins  
CC and a source of carbohydrates, especially a mixture of leguminous plant  
CC or of a cooked oleaginous plant and of a cooked or roasted cereal source  
CC e.g. a mixture of soya or cooked beans and of cooked or roasted wheat or  
CC rice.  
XX  
SQ Sequence 963 AA;  
Query Match 29.6%; Score 1405.5; DB 20; Length 963;  
Best Local Similarity 34.2%; Pred. No. 8.3e-110;  
Matches 311; Conservative 161; Mismatches 292; Indels 145; Gaps 24;  
Qy 38 GYVRSVGVDSNRQVLTAKDLIKPSSUYADPIKSLNLHVSLETSELRIRI----TDSS 93  
Db 72 GYEL--VNVTTAKLTGLKLNEATNIYGYDFYDNLNLSVEYQSDRLNVHIEPYDTDNV 129  
Qy 94 QORWEIPETVIPRAGNHSRRSTEDCGNSPENNFADPSDVLFTLHNTTFFGFSYSR 153  
Db 130 ---FILPESLVAK-----PSADGDGKIE-SFHEGSSDLVFE-YSSKNFGFEILR 174  
Qy 154 RSSGDILFDTSPDSSDSNTYFIKQDFLQLSALPENRSLNYGIGEHYKRSFRLIPGEM 213  
Db 175 KSTGKSIFSTIGNP-----LVFSNQIFQNTSLPKDHF-ITGLGE-SIHGPRNEPGIVK 226  
Qy 214 TLWNADIGSENP-DYNLYGSHPPYMDVRGSKNEEAGTHGVLLNSNGMDVKEYGHRIT 272  
Db 227 TIYANDIA--NPIDGNIYGVHPPYIDQR-----FTNATHGVYMTSAIQEAVAVNESLT 279  
Qy 273 YNVIGGVIDLVEFAGSPPEMVMNOYTELIGRPAPMPYMSFGHQCRYGYKXVSDLEYVVD 332  
Db 280 WRALSGVDLFFSGPKPKDVIQYVKEVGLPTFPYWALGYHQCRWGYDIEELDEVE 339  
Qy 333 GYAKAGIPELVNMTDIDYMDGYKDFTLDPVNPFPEDKMQSPVDTLHKNGQKYVLIIDPGIG 392  
Db 340 NPKKPDIPLETIWSIDYMDSYKDFTNDRPHRYPLEKYQQFLDKLHNNQRYVPIIDAAIY 399  
Qy 393 V-----DSSYGTYNRCMEADVFIKR-NGEPYLGEVWPQKVPYDFDLPNPAATFWSNEI 444  
Db 400 VNPENATDNDYDVHYGNETDVLKNPDGSLYIGAVMPGYTVFPDFLSENITQYTKTVK 459  
Qy 445 KMFOEILPLDGLWIDMELSNF----- 466  
Db 460 KDWYQOIKFDPGLWIDMELSNF----- 466



```

QY 467 ----ITPLSSGSSLD-----PPYKINNSGDKRPINN 495
Db 520 EYASFTSLAASSTDESSVSSTASIDSLNTPARCKGNINYPPIAINNDQGDHDLAT 579
QY 496 KTVPATSIHFNISYDAHNLVGLLEAKATHQAVVDI-TGKRPFILSRSTFVSSGKYTAH 554
Db 580 HAVSPNATHQDGTLEYDVHNLVGLYLETNATFEALLE:QPNKRPFILSRSTFVSSGKYTAH 639
QY 555 WTGDNAAKWEDLAYSIFGILNFGLPFGIPMGADICGFSHDTTELCRMIQLGNFYPPAR 614
Db 640 WGGDNYSQFRAYSIAQAFSGUSGIPFFGADVCGFNGSDYELCSRMMQLGSGFFPYR 699
QY 615 DHSSLGTAARQELYLWDSVASSARKVLGRMLLPHLYTLMEYAHVSGNPIARPLFFSPQ 674
Db 700 NHNIGALISQEPYVESVTEATKTSQMLRYLLLPYVYLLHEHTTGPILRAFAMQFPE 759
QY 675 DTKYEYDSQFLGKSMVSPALQKQAVADAYFP-AGN-----WFDLNYFSAVGSDS 727
Db 760 NKNYSTVDTPQFVGDALVVTPLVLEQGVDTVKGTFFGSGNEEYVYDWTHEKQNT----D 815
QY 728 GKHYRLDTPADHVNHVREGSIVAMQGEALTTDRARXTEYQLLVASRLNIGSEFLDD 787
Db 816 GKNETLOAPLGHIPHIRGCHILTPQEPAYTTTESRQNPWGLIADKQKAEKGSYSD 875
QY 788 GENLRMGAGGGRDWTLYKPCYVTKSVLVRSEVNVPEYAS-KMKWSIGKVTFFVGPNV 846
Db 876 GESYEV-----EBSLFVNF-----IASDNTLLSTSYGEYVEQPLAN-TILGVEN- 920
QY 847 ENVTYEV 855
Db 921 ---KPKVK 926

RESULT 12
AAW74090.
ID AAW74090 standard; Protein; 1827 AA.
XX
AC AAW74090;
XX
DT 04-MAY-1999 (first entry)
XX
DE Human hSI protein sequence.
XX
KW Gastro-intestinal transport receptor; binding protein; hSI; HPT1;
KW D2H; hPEP1; human; GI tract receptor; sucrose-isomaltase complex;
KW intestinal peptide-associated transporter; hypertension; diabetes;
KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
KW therapeutic agent delivery; therapy; probe.
XX
OS Homo sapiens.
XX
PN MO9851325-A2.
XX
PD 19-NOV-1998.
XX
PF 15-MAY-1998; 98WO-US10088.
XX
PR 15-MAY-1997; 97US-0046595.
XX
PA (CYTO-) CYTOGEN CORP.
XX
FA (ELAN-) ELAN CORP PLC.
XX
PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin IJ;
PI Omahony DJ, Patterson CA, Singleton J;
XX
DR WPI; 1999-009568/01.
XX
PT New proteins that bind specifically to receptors in the
PT gastro-intestinal tract and related nucleic acid - chimaerae and
PT antibodies, used to deliver therapeutic or diagnostic agents to, or
PT through, the gastrointestinal tract, e.g. insulin or leuprolide
XX
PS Disclosure; Fig 3; 294pp; English.

```

```

XX This sequence is the human hSI protein. The invention relates to
CC purified proteins (I) that bind specifically to at least one of the
CC gastro-intestinal (GI) tract receptors human intestinal
CC peptide-associated transporter (HPT1), hPEP1, D2H and human
CC sucrose-isomaltase complex (hSI). (I) provide active transport of
CC therapeutic agents through human and animal GI tissue (into the blood)
CC for in vivo delivery, particularly for treatment or prevention
CC of hypertension, diabetes, osteoporosis, haemophilia, anaemia, cancer,
CC migraine, or angina pectoris. Specifically they are used to deliver
CC insulin or leuprolide, but many other suitable therapeutic agents are
CC disclosed, including genes or inhibitory nucleic acid, imaging agents and
CC antigens. (I) may also provide targeting to the GI tract. Other uses of
CC (I) are: (i) to determine the level of specified receptors in a sample
CC (in a binding assay); and (ii) to screen for molecules that bind (I).
CC immunogenic analogues or derivatives of (I) are used to raise antibodies
CC and in immunoassays. The antibodies are used to locate, detect and
CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
CC etc., also for peptide purification and immobilisation.
XX
SQ Sequence 1827 AA;
Query Match: 28.6%; Score 1360; DB 20; Length 1827;
Best Local Similarity 38.2%; Pred. No. 1.7e-105;
Matches 301; Conservative 138; Mismatches 258; Indels 90; Gaps 21;
QY 34 VVGYGYYVR-----SVGVDSNRQVLTAKLDIKPSSVYAPDIKSLNLHVLSLETSELRIR 88
Db 109 VDNHGYNQDMTTSIGVE-----AKLNRPSTPLFGNDINSVLFTTQNTPNRFRK 161
QY 89 ITDSSQORWEIPETVIPRAGNHSRFRFSTEEDGGNSPENNFLADPSSDLVFTLH-NTTFF 147
Db 162 ITDPNRRAYEYVQV-----KEFTGPTVSDTLVDVKAQNPFF 199
QY 148 GFSVSRSSGDI LFTSPDSDSNITYFIKQOFLQLSALPENRNLXGIGHTKRSFL 207
Db 200 SIQVIRKSNKGLTFTSIGP-----LVYSQYLOISARLPSD--YIYGIGBOVHKRFRH 251
QY 208 -IPGETMTLWADIGSENPVNLYGSHPFYMDVROSKGNEEAGTTGVLLNLSNMDVYK 266
Db 252 DLSWKTPIFTTRDQLPGNNNNLYGHQTFMCI-----EDTSGKSGFVFLMNSNAMEIFI 306
QY 267 EGHRI-TYNVIGGVADLVYFAGSPSEVMYQYTELIGRPAMPYMSFGPHOCRYKYNVS 325
Db 307 QTPPIVTVYRGILDFYLLGDTPEVQVQVQQLVGLPAMPAYWNLGFLSRWYKSLD 366
QY 326 DLEYVDVYAKAGIPLEVWMTDIDMVGKOPTLDPVNPFPEDKMQSFVDTLHKNQKYL 385
Db 367 VYKVVRRNREAGIPEDTQVTDIDYMDKQDFTYDQVAF--NGLPQVQDLHDHQKYVI 424
QY 386 ILDRGIGV-----DSSYGTYNRGMADYPIKRN--GEPVLGEVWPKVYFPDFLPAAT 438
Db 425 ILDPALISGRRANGTYATYERGNTQHWINESDGTPIIGVWPGLTVPYDFTWPCID 484
QY 439 FWSNEIKMFOEILPLDGLWDMNLSNFTSPLS--SGSSLDPPYKINNSGD--KRPN 494
Db 485 WMANECISFHOEVQYDGLWDMNEVSSFIQGSTKCNVKNLYNPPF---TPDILDKLWY 540
QY 495 NKTVPATSI-HFGNISEVDARNLYGLLEAKATHQAVVDI-TGKRPFILSRSTFVSSGKYT 552
Db 541 SKTICMDAVQNWG--KQYDVHSLYGSMAIATEAQAVQVFNKRSFILTRSTFAGSGRHA 598
QY 553 AHWTDGNAKWEDLAYSIPGILNFGLPFGIPMGADICGFSHDTTELCRMIQLGCAFPF 612
Db 599 AHWLGDNTASWQMEWSITGMLFSLFGIPLVGADICGFVAETTELCRRMMQLGAPFP 658
QY 613 ARDHSSLGTAARQELYLWDS---VASSARKVLGRMLLPHLYTLMEYAHVSGNPIARPLF 669
Db 659 SRNHSNDGVEHQDPAFFQNSLVKSSRQYLTIYTLPLFLYLYKAHVGETVAPV 718
QY 670 FSPQDQTKYEYDSQFLGKSMVSPALQKQAVADAYFPAGNWFDFLNYFSAVGSDS- 728
Db 719 HEFYDNTSNWIEDTFLMGPAALLITPVLKQAGADTVSAVIPDAIWDY-----ESGA 769

```

QY 729 -----KHVRLDTPADHVNHVREGSVIVAMQGEALTTDRDKTPYQOLLVVASRLENISGE 782  
 DB 770 KRPWRKORVDVYLPADKIGLHRLGGYIIPIQSPDVTITASRKNPLGLIVALGENNTAKGD 829

QY 783 LFLDDGE 789  
 DB 830 FFWDDGE 836

RESULT 13  
 ID AAU09028  
 AC AAU09028  
 XX 18-DEC-2001 (first entry)  
 DT Human sucrase isomaltase, SI.  
 XX Human, sucrase isomaltase; SI; stomach cancer; colorectal cancer;  
 KW oesophageal cancer; cytostatic; metastasis; CDX1; CDX2; tumour.  
 XX Homo sapiens.  
 OS WO200173:33-A1.  
 PN 04-OCT-2001.  
 PD 27-MAR-2001; 2001WO-US09918.  
 PF 27-MAR-2000; 2000US-192229P.  
 PR (UYJE-) UNIV JEFFERSON THOMAS.  
 PA Waldman SA, Park J, Schulz S;  
 XX WPI; 2001-6:1641/70.  
 DR N-PSDB; AAS14674.  
 XX In vitro screening for specific gastrointestinal cancer cells, useful  
 PT for diagnosis, by detecting expression of the markers SI, CDX1 or CDX2  
 PT -  
 XX Disclosure; Page 114-119; 119pp; English.

CC The invention relates to in vitro screening of metastatic colorectal  
 CC cancer cells or primary and/or metastatic stomach or oesophageal cancer  
 CC cells by testing cells in extra-intestinal tissues and/or body fluids for  
 CC expression of SI (sucrase isomaltase), CDX1 or CDX2 (transcription  
 CC factors). Expression of these markers indicates possible presence of the  
 CC specified cancer cells. The method is used to diagnose (or monitor)  
 CC metastatic colorectal cancer or primary and/or metastatic stomach or  
 CC oesophageal cancer cells, also to confirm identification of such cells.  
 CC These cancers can be treated by administration of an SI ligand and  
 CC (optionally conjugated) cytostatic agent or radioimaged by administering  
 CC a conjugate of the SI ligand and detectable agent. The present sequence  
 CC represents human SI.  
 XX Sequence 1827 AA;

Query Match 28.6%; Score 1360; DB 22; Length 1827;  
 Best Local Similarity 38.2%; Pred. No. 1.7e-105;  
 Matches 301; Conservative 138; Mismatches 258; Indels 90; Gaps 21;

QY 34 VVGGVYVR-----SVGVDSNRQVLTAKLDIKSSVYAPDKSLNHLVSLTSELRIR 88  
 DB 109 VDNHGVYVQDMTTTISIGVE-----AKLARIPSTFLFGNDINSVLFTTQNTQNRFRFX 161  
 QY 89 ITDSSQORWEIPETVIPRAGNHSRRSTEDDGCNSPENNFLADPSSDLVFTLH-NTTPF 147  
 DB 162 ITDPNNRRYEVPHQV-----KFTGTPTVSDTLIDYKVAQNPF 199

QY 148 GFSVSRSSGDI LDTSPDSSDNTYFIFKDOQLQLSALLPENRSLNYGIGETKRFRL 207  
 DB 200 SIQVIRKSNKGLT LDTSGP-----LVYSDOYLQISARLPD--YIYGEGEVHKRFRH 251  
 QY 208 -IPGETWTLNADTGSBNPDVNLVYGSHPFYMDVRGSKGNEEAGTHGVLLNSNGMDVKY 266  
 DB 252 DLSWKTWPIFTRDQLPGDNNNLYGHOTFFMCI-----EDTSKSGFVFLMNSNAXEIFI 306  
 QY 267 EGHRI-TYNVIGGVIDLVPFAGPSPVPMQMYTELIGRPAHPMPYMSFGHCRCRYGKNVS 325  
 DB 307 QPTPIVTVRTGGILDFYLLGDTPEQVVQYQQLVGLPAMPAYWYNIQFQLSRWNYKSLD 366  
 QY 326 DLEYVVDGYAKAGIPLEVMTDIDYMDGYDFTLDPVNFPEDEKMQSFVDTLHKMGQKYVL 385  
 DB 367 VVKEVVRNREAGIPFDQVTDIDYMDKKDFTYDQVAF--NGLPQFVQDLHDHGQKYVI 424  
 QY 386 ILDPGIGV-----DSSYGTNRMGEADVFKRN--GEPYLGGEVWPGKYVPDFLNPAAT 438  
 DB 425 ILDPAISIGRRANGTATYERGNTOHVWYNESDGSTPIIGEVMPGLTVYPDFTNPCID 484  
 QY 439 FWSNEIKMFQELPLDGLWIDMNSLNFITSPLS--SGSSLDDPPYKINNSGD--KRPIN 494  
 DB 485 WNAECISIFHOEVQVYDGLWIDMNEVSSFIQSGTQGVNKLNYPPF---TPDILDKLKY 540  
 QY 495 NKTVPATSI-HFGNISEYDAHNLYGLLEAKATHQAVVDI-TGRKPPFILSRSTFVSSGKYT 552  
 DB 541 SKTICMDAVQNWG--KOYDVHSLYGSMAIATEQAVQKVPFNKRSFILTRSTFAGSGRHA 598  
 QY 553 AHWYGDNAKVEDLAYSIPIGLNFGIGIPMGVADICGPHSDTTEELCRHWIQLGAYYFP 612  
 DB 599 AHWLGDNTASWEQMEWSITGMLEFSLFGIPLVGADICGFVAETTEELCRMMQLGAYYFP 658  
 QY 613 ARDHSSLGTAQELYLWDS---VASSARKVLGLRMLLPHLYTLIMYEAHVSGNPIARPLF 669  
 DB 659 SRNHSNDCYEHQDPAPFGQNSLLVKSRSQYLTIHYLLPFLYTLFYKAHVGETVARPVL 718  
 QY 670 FSPQDTKTYEIDSOFLIGKSIWVSPALKQGAVADAYFPAGNWFDFNYSFAVGDSG- 728  
 DB 719 HEFYEDTNSWIEDTEFLWGPALLITPVLKQGADTVSAYIPDAIWYD-----BSGA 769  
 QY 729 -----KHVRLDTPADHVNHVREGSVIVAMQGEALTTDRDKTPYQOLLVVASRLENISGE 782  
 DB 770 KRPWRKORVDVYLPADKIGLHRLGGYIIPIQSPDVTITASRKNPLGLIVALGENNTAKGD 829

QY 783 LFLDDGE 789  
 DB 830 FFWDDGE 836

RESULT 14  
 ID AEG14994  
 AC AEG14994  
 XX 18-FEB-2002 (first entry)  
 DT Novel human diagnostic protein #14985.  
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 OS Homo sapiens.  
 XX WO200175067-A2.  
 XX 11-OCT-2001.  
 XX 30-MAR-2001; 2001WO-US08631.  
 XX 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX

(HYSE-) HYSEQ INC.

Dmanac RT, Liu C, Tang YT;

WPI; 2001-639362/73.

DR N-PSDB; AAS79181.

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity

Claim 20; SEQ ID No 45353; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 1829 AA;

Query Match 28.6%; Score 1360; DB 22; Length 1829;  
Best Local Similarity 38.2%; Pred. No. 1.7e-105;  
Matches 301; Conservative 138; Mismatches 258; Indels 90; Gaps 21;

Qy 34 VVGYYVVR-----SVGVDSNRQVLTAKDLIKPSSVYAPDIKSLNLSVLETSELRIR 88  
Db 111 VDNHGYVQDMMTTTSIGVE-----AKLNRPSTPLFGNDINSVLETTQNTNRRFEK 163  
Qy 89 ITDSQQRWEIPEVI PRAGNHSRRFSTEEDGGNSPENNFLADPSSDLVFTLH-NTTPF 147  
Db 164 ITDENRRRYEVPHQV-----KEFTGPTVSDTLVDVKAQNP 201  
Qy 148 GFSVSRSSGDLFDTSPOSSDSNTYFIFKQQLQSSALPENRSLNYGIGEHTRKSRFL 207  
Db 202 SIQVIRKNGKTLFTSISG-----LVYSQYLOISARLPD--YIYGEGVQVHKBFRH 253  
Qy 208 -IPGETMTLNADIGSEPNVDNLYGSHFFYMDVRGSKNEEAGTHGVLLNSNGMDVKY 266  
Db 254 DLSWKTWPIFTRDQLPGDNNNLYGHQTFMCI-----EDTSKSGFVFLMNSAMEIFI 308  
Qy 267 EGHRI-TYNVLGGVLDLVFAGPPEMNMNYTELIGRAPMPYMSRGEHCYKXNS 325  
Db 309 QFTPIVTRVGGILDVILLGDTPEQVQVQQLVGLPAMPAYNVLGFSQLSRNYSKLD 369  
Qy 326 DLEYVVDGYAKAGIPLVEMWTDIDYMDGYKFTLDPVNFPFCKMQSFVDTLHKQKQKYL 385  
Db 369 VVKEVVRNREAGIPFDQVTDIDYMDKDKFTYDQAF--NGLPQFQDLHDHGQKVI 426  
Qy 386 ILDPGIGV-----DSSYGTNRGMEADVIFKRN--GEPYLGEVMPGKYVPDFLNPAAAT 438  
Db 427 ILDPAISICRRANGTYYATYERGTQHWINESDGSSTPIIGEVWFGTLVYPDFTPNCID 486  
Qy 439 FWSKEIKMFOEILPLDGLWIDMNSLNFITSPLS--SGSSLDSDPYKLNNSGD--KRPIN 494  
Db 487 WMANECFHOEVQVQDGLWIDMNEVSSFOGSKGNCVKNLNYPPF-----TPDILDKLMY 542

QY 495 NKTVPATSI-HFGNISEYDAHNLYGLLEAKATHOAVVDI-TCKRPFILSRSTFVSGKYT 552  
Db 543 SKTICMDAVQNWG--KOYDVHSLYGYSMATIEQAQVQVFPNKRSEFILTRSTFAGSGRHA 600  
QY 553 AHWTDGNAKWEDLAYSIPIGILNFGILFGIPMVGADICGFSHDTTEELCRRWTQLGAFYPP 612  
Db 601 AHWLDGNTASWEQMEWSITGMLFSLFGILPLVGADICGFAETTEELCRRWMLGAFYPP 660  
QY 613 ARDHSLSGTARQELYLWDS---VASSARKVLGLRMLLPHLYTLMYEAVHVSNGNPJAPL 669  
Db 661 SRNHNSDGYEHQDPAPFFQNSLLVKSSRQYLTRYTLFLFYLYTKAHVFGTAVRVL 720  
QY 670 FSPQDITKYBIDSOFLICKSIMVSPALKQGVAVDVFAGNWFEDLNFYSFAVGDSG- 728  
Db 721 HEFYEDTNSWIEDTEFLWGPALLITPVLKQGDVTSAYIPDAIWYD-----ESGA 771  
QY 729 -----KHVRLDTPADHNVHVRGSIAMQGEALTTDRDKTPYOLLVVASRLNISE 782  
Db 772 KRPWRKQRYDMYLPADKIGLHLRGYIPIQEPDVTTTASRKNPLGLIVALCENNTAKGD 831  
QY 783 LFLDDGE 789  
Db 832 FFWDDGE 838

## RESULT 15

AAW15191  
ID AAW15191 standard; Protein; 985 AA.

AC AAW15191;

DT 18-JUN-1997 (first entry)

DE Aspergillus oryzae alpha-glucosidase.

XX Fungus; fungal; enhancer element; promoter; recombinant protein.

XX Aspergillus oryzae.

PN JP09009968-A.

PD 14-JAN-1997.

PF 29-JUN-1995; 95JP-0163579.

PR 29-JUN-1995; 95JP-0163579.

PA (KOKU-) KOKUZEI CHO CHOHAN.

PA (OZEK-) OZEKI KK.

XX WPI; 1997-126425/12.

DR N-PSDB; AAT65017.

XX Fungal DNA enhancer element - used to transform other host fungus e.g. Aspergillus oryzae, to produce large quantities of a gene product

PS Disclosure; Page 14-19; 25pp; Japanese.

XX The promoter region of the alpha-glucosidase (agda) gene of Aspergillus oryzae contains two novel enhancer elements: one (designated "enhancer-B") corresponds to the consensus sequence CCGNATTTA and the other (designated "enhancer-C") is of sequence CCAATCAGCGT. By inserting at least one of the enhancer elements into a promoter region which is functional in fungi, the activity of the promoter is enhanced. Using such improved promoters, a gene of interest can be expressed efficiently in transformed fungi. The present sequence is encoded by the agda gene.

SQ Sequence 985 AA;

Query Match 28.6%; Score 1357; DB 18; Length 985;



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2003, 10:20:37 : Search time 2569 Seconds

(without alignments)  
274.360 Million cell updates/sec

Title: US-10-043-418-5

Perfect score: 29

Sequence: 1 cggtagattgacagatccaaggtaga 29

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estlin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pin:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_man:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rod:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrl:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	25.8	89.0	538	12	BM375479
C 2	25.8	89.0	628	13	B0739877
C 3	22.6	77.9	468	14	CA724231
C 4	21.6	74.5	749	28	AZ175770

5	21	72.4	319	9	AU162340
C 6	21	72.4	554	10	BF943271
7	21	72.4	635	14	CA764226
C 8	21	72.4	646	10	BE262412
9	21	72.4	671	28	AQ955967
C 10	21	72.4	688	10	BE260349
C 11	21	72.4	845	10	BE312763
C 12	21	72.4	900	13	B0859266
C 13	21	72.4	905	13	B0159643
C 14	21	72.4	908	10	BF206541
C 15	21	72.4	932	13	B0845072
C 16	21	72.4	1100	12	BI255751
C 17	21	72.4	1181	10	BF038977
C 18	20.6	71.0	409	28	AZ098515
19	20.6	71.0	495	12	B018010
C 20	20.6	71.0	594	28	AQ924809
C 21	20.6	71.0	601	28	AZ059272
C 22	20.6	71.0	646	28	AZ291074
23	20.6	71.0	713	12	B020716
24	20.6	71.0	1131	14	CD498216
25	20.4	70.3	965	13	BX404948
26	20.4	70.3	1003	9	AL568184
C 27	20.2	69.7	262	10	BE020298
C 28	20.2	69.7	476	10	BF775363
C 29	20.2	69.7	721	12	B0044415
30	20.2	69.7	748	10	BE616952
C 31	20.2	69.7	753	12	BM951204
C 32	20	69.0	214	28	AZ112217
33	20	69.0	471	14	CB212812
34	20	69.0	552	10	BF943732
35	20	69.0	575	12	B0111443
C 36	20	69.0	576	28	AZ087732
C 37	20	69.0	655	9	AL871784
C 38	20	69.0	792	28	BZ061211
C 39	20	69.0	973	12	BI950841
C 40	20	69.0	994	12	BI196323
41	19.8	68.3	322	13	BY175032
42	19.8	68.3	847	9	AL565539
43	19.8	68.3	1201	13	BX423745
C 44	19.6	67.6	268	14	W76119
C 45	19.6	67.6	275	10	BB158774

ALIGNMENTS

RESULT 1	BM375479/c	BM375479	538 bp	mRNA	linear	EST 23-JUL-2002
LOCUS	EBem06_SQ003_H15_R	embryo, 21 DPA, no treatment, cv Optic, EBem06				
DEFINITION	Hordeum vulgare subsp. vulgare	cdna clone EBem06_SQ003_H15 5', mRNA				
ACCESSION	BM375479					
VERSION	BM375479.2	GI:21935437				
KEYWORDS	EST.					
SOURCE	Hordeum vulgare subsp. vulgare					
ORGANISM	Hordeum vulgare subsp. vulgare					
REFERENCE	1 (bases 1 to 538)					
AUTHORS	Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardle, L., Ramsay, L., Machray, G., Marshall, D.F.M. and Waugh, R.					
TITLE	Development of Barley Transcriptome Resources					
JOURNAL	Unpublished					
COMMENT	On Jan 10, 2002 this sequence version replaced gi:18118869. Contact: Waugh R, Marshall DF Genome Dynamics/Computational Biology Scottish Crop Research Institute Invergowrie, Dundee, DD2 5DA, Scotland, UK Tel: 00 44 1382 562731 Fax: 00 44 1382 562426 Email: est@scri.sari.ac.uk					

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: LLAM13456 row: 0 column: 13  
High quality sequence stop: 669.

FEATURES  
source  
1. .905  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6139620"  
/tissue\_type="retinoblastoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_67"  
/note="Organ: eye; Vector: pCMV-SPORT6; Site: 1: NotI; Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Library constructed by Life Technologies."

BASE COUNT 203 a 279 c 259 g 164 t  
ORIGIN

Query Match 72.4%; Score 21; DB 13; Length 905;  
Best Local Similarity 82.8%; Pred. No. 5.6e+02;  
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGGTGAAGTTGACAGGATCCAAAGGTGAAG 29  
||||| ||||| ||||| ||||| |||||  
DB 656 CGGTGAAGATGACAGGTCCTCATGAGGAAG 628

RESULT 14  
BF206541/c  
LOCUS  
DEFINITION  
601869770P1 NIH\_MGC\_19 Homo sapiens cDNA clone IMAGE:4100105',  
mRNA sequence.  
ACCESSION  
BF206541  
VERSION  
BF206541.1 GI:11100127  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
1 (bases 1 to 908)  
AUTHORS  
NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL  
Unpublished  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [image.llnl.gov](http://image.llnl.gov)  
Plate: LLCW969 row: k column: 18  
High quality sequence stop: 602.

FEATURES  
source  
1. .908  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4100105"  
/tissue\_type="neuroblastoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_19"  
/note="Organ: brain; Vector: pOTB7; Site: 1: XhoI; Site: 2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

BASE COUNT 159 a 293 c 242 g 206 t  
ORIGIN

Query Match 72.4%; Score 21; DB 13; Length 900;  
Best Local Similarity 82.8%; Pred. No. 5.6e+02;  
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGGTGAAGTTGACAGGATCCAAAGGTGAAG 29  
||||| ||||| ||||| ||||| |||||  
DB 170 CGGTGAAGATGACAGGTCCTCATGAGGAAG 142

RESULT 13  
BU159643/c  
LOCUS  
DEFINITION  
BU159643  
ACCESSION  
BU159643  
VERSION  
BU159643.1 GI:22673553  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
1 (bases 1 to 905)  
AUTHORS  
NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL  
Unpublished  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: LLAM13456 row: 0 column: 13  
High quality sequence stop: 669.

FEATURES  
source  
1. .905  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6139620"  
/tissue\_type="retinoblastoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_67"  
/note="Organ: eye; Vector: pCMV-SPORT6; Site: 1: NotI; Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Library constructed by Life Technologies."

BASE COUNT 203 a 279 c 259 g 164 t  
ORIGIN

Query Match 72.4%; Score 21; DB 13; Length 905;  
Best Local Similarity 82.8%; Pred. No. 5.6e+02;  
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGGTGAAGTTCACAGGATCCAAAGGTGAAG 29  
||||| ||||| ||||| ||||| |||||  
DB 656 CGGTGAAGATGACAGGCTCCATGAGGAAG 628

RESULT 14  
BF206541/c  
LOCUS  
DEFINITION  
601869770P1 NIH\_MGC\_19 Homo sapiens cDNA clone IMAGE:4100105',  
mRNA sequence.  
ACCESSION  
BF206541  
VERSION  
BF206541.1 GI:11100127  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1. (bases 1 to 908)  
AUTHORS  
NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL  
Unpublished  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [image.llnl.gov](http://image.llnl.gov)  
Plate: LLCW969 row: k column: 18  
High quality sequence stop: 602.

FEATURES  
source  
1. .908  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4100105"  
/tissue\_type="neuroblastoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_19"  
/note="Organ: brain; Vector: pOTB7; Site: 1: XhoI; Site: 2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

BASE COUNT 159 a 293 c 242 g 206 t  
ORIGIN

Query Match 72.4%; Score 21; DB 13; Length 900;  
Best Local Similarity 82.8%; Pred. No. 5.6e+02;  
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGGTGAAGTTCACAGGATCCAAAGGTGAAG 29  
||||| ||||| ||||| ||||| |||||  
DB 170 CGGTGAAGATGACAGGCTCCATGAGGAAG 142

RESULT 13  
BU159643/c  
LOCUS  
DEFINITION  
BU159643 NIH\_MGC\_67 Homo sapiens cDNA clone IMAGE:6139620  
s', mRNA sequence.  
ACCESSION  
BU159643  
VERSION  
BU159643.1 GI:22673553  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1. (bases 1 to 905)  
AUTHORS  
NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL  
Unpublished  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-f@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation

```
BASE COUNT      223 a      249 c      246 g      190 t
ORIGIN
Query Match      72.4%; Score 21; DB 10; Length 908;
Best Local Similarity 82.8%; Pred. No. 5.6e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1  CGGTGAAGTTGACAGGATCCAGGTGAAG 29
      |||||  |||||  |||||  |||||  |||||
Eb      464  CGGTGAAGATGACAGGGTCCATGAGGAAG 436

RESULT 15
BU845072/c
LOCUS
DEFINITION      BU845072      932 bp      mRNA      linear      EST 16-OCT-2002
IMAGE:6578606 5', mRNA_sequence.
ACCESSION      BU845072
VERSION
KEYWORDS
SOURCE      BU845072.1 GI:24029513
ORGANISM      Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 932)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: ARCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCN2780 row: j column: 14
High quality sequence stop: 734.
FEATURES
Location/Qualifiers
1..932
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6578606"
/tissue_type="teratocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_109"
/notes="Organ: ovary; Vector: pOT87; Site 1: EcoRI; Site 2:
XhoI; CDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
BASE COUNT      169 a      302 c      247 g      211 t      3 others
ORIGIN
Query Match      72.4%; Score 21; DB 13; Length 932;
Best Local Similarity 82.8%; Pred. No. 5.7e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1  CGGTGAAGTTGACAGGATCCAGGTGAAG 29
      |||||  |||||  |||||  |||||  |||||
Eb      181  CGGTGAAGATGACAGGGTCCATGAGGAAG 153

Search completed: October 27, 2003, 11:54:52
Job time : 2577 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2003, 10:20:37 ; Search time 1420 Seconds

(without alignments)  
54.769 Million cell updates/sec

Title: US-10-043-418-5

Perfect score: 29

Sequence: 1 cggtaggtgacagatccaaggtgaag 29

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1792395 seqs, 1340900451 residues

Total number of hits satisfying chosen parameters: 3584790

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/prodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/prodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/prodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/prodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/prodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/prodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/prodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/prodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/prodata/1/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/prodata/1/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/prodata/1/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/prodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/prodata/1/pubpna/US10A\_PUBCOMB.seq:\*
- 14: /cgn2\_6/prodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/prodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 16: /cgn2\_6/prodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 17: /cgn2\_6/prodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	100.0	29	13	US-10-043-418-5
2	21.6	74.5	3186778	13	US-10-027-632-174961
C 3	21	72.4	215	11	US-09-844-653-44
C 4	21	72.4	5659	14	US-10-225-567A-521
C 5	21	72.4	48841	11	US-09-844-653-32
6	20	69.0	944	9	US-09-987-190-13
7	20	69.0	1476	12	US-10-032-585-6687
8	20	69.0	1750	9	US-09-987-190-7
9	19.4	66.9	9025608	14	US-10-156-761-1
C 10	19.2	66.2	526	13	US-10-027-632-5351
C 11	19	65.5	487	11	US-09-918-995-1499
C 12	19	65.5	106664	14	US-10-175-523-97
C 13	18.8	64.8	2656	14	US-10-128-714-7
14	18.8	64.8	2807	14	US-10-128-714-5007
C 15	18.6	64.1	713	14	US-10-016-253-13
16	18.6	64.1	1046	13	US-10-027-632-9163

17	18.4	63.4	60	12	US-09-908-975-31331	Sequence 31331, A
18	18.4	63.4	80	11	US-09-992-331-5	Sequence 5, Appli
19	18.4	63.4	80	15	US-10-262-313-5	Sequence 5, Appli
C 20	18.4	63.4	576	9	US-09-864-761-16455	Sequence 16455, A
C 21	18.4	63.4	748	13	US-10-027-632-141845	Sequence 141845,
C 22	18.4	63.4	823	11	US-09-791-932-7	Sequence 54, Appli
C 23	18.4	63.4	1464	10	US-09-771-161A-54	Sequence 607, App
C 24	18.4	63.4	1506	10	US-09-938-842A-607	Sequence 1, Appli
C 25	18.4	63.4	1508	14	US-10-146-835-1	Sequence 3, Appli
C 26	18.4	63.4	1541	14	US-10-146-835-3	Sequence 53, Appli
C 27	18.4	63.4	2518	10	US-09-771-161A-53	Sequence 366, App
28	18.4	63.4	2988	12	US-10-354-358-73	Sequence 366, App
C 29	18.4	63.4	2988	14	US-10-171-581-366	Sequence 602, App
C 30	18.4	63.4	14414	14	US-10-225-567A-602	Sequence 739, App
C 31	18.4	63.4	35178	12	US-10-017-161-739	Sequence 154, App
C 32	18.4	63.4	302250	10	US-09-962-832-154	Sequence 3634, Ap
33	18.2	62.8	510	12	US-09-814-353-3634	Sequence 9948, Ap
34	18.2	62.8	510	12	US-09-814-353-9948	Sequence 14204, A
35	18.2	62.8	544	12	US-09-814-353-14204	Sequence 288143,
C 36	18.2	62.8	558	13	US-10-027-632-288143	Sequence 288144,
C 37	18.2	62.8	558	13	US-10-027-632-288144	Sequence 288145,
C 38	18.2	62.8	558	13	US-10-027-632-288145	Sequence 1459, Ap
39	18.2	62.8	617	12	US-09-814-353-1459	Sequence 7819, Ap
C 40	18.2	62.8	617	12	US-09-814-353-7819	Sequence 288142,
C 41	18.2	62.8	661	13	US-10-027-632-288142	Sequence 16332, A
42	18.2	62.8	671	12	US-09-814-353-16332	Sequence 19856, A
43	18.2	62.8	687	12	US-09-814-353-19856	Sequence 729, App
C 44	18.2	62.8	5671	11	US-09-764-872-729	Sequence 190, App
C 45	18	62.1	556	12	US-10-205-219-190	

#### ALIGNMENTS

RESULT 1  
US-10-043-418-5  
; Sequence 5, Application US/10043418  
; Publication No. US20020184662A1  
; GENERAL INFORMATION:  
; APPLICANT: Henson, Cynthia A.  
; APPLICANT: Muslin, Elizabeth H.  
; APPLICANT: Clark, Suzanne E.  
; TITLE OF INVENTION: Modified barley alpha-glucosidase  
; FILE REFERENCE: 960296.97486  
; CURRENT APPLICATION NUMBER: US/10/043.418  
; CURRENT FILING DATE: 2002-06-25  
; PRIOR APPLICATION NUMBER: 60/260,787  
; PRIOR FILING DATE: 2001-01-10  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 29  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:synthetic  
; OTHER INFORMATION: oligonucleotide  
US-10-043-418-5

Query Match . 100.0%; Score 29; DB 13; Length 29;  
Best Local Similarity 100.0%; Pred. No. 0.0054;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGGTGAAGTTCACAGGATCCCAAGGTGAAG 29  
|||||  
Db 1 CGGTGAAGTTCACAGGATCCCAAGGTGAAG 29

RESULT 2  
US-10-027-632-174961  
; Sequence 174961, Application US/10027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.



; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 174961  
; LENGTH: 3186778  
; TYPE: DNA  
; ORGANISM: Human  
; NAME/KEY: misc feature  
; LOCATION: (1)..-(3186778)  
; OTHER INFORMATION: n = A,T,C or G  
US-10-027-632-174961

Query Match 74.5%; Score 21.6; DB 13; Length 3186778;  
Best Local Similarity 85.7%; Pred. No. 15;  
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 2 CGGTGAAGTTGACAGGATCCCAAGGTGAAG 29  
Db 2065153 GGGGAATTCACAGCATCCCAAGGTGAAG 2065180

RESULT 3  
US-09-844-653-44/c  
; Sequence 44, Application US/09844653  
; Publication No. US20030054347A1  
; GENERAL INFORMATION:  
; APPLICANT: Richards, Julia  
; APPLICANT: Rozsa, Frank  
; TITLE OF INVENTION: Detecting and Treating Eye Disease  
; FILE REFERENCE: UM-06105  
; CURRENT APPLICATION NUMBER: US/09/844,653  
; CURRENT FILING DATE: 2001-04-27  
; NUMBER OF SEQ ID NOS: 173  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 44  
; LENGTH: 215  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-844-653-44

Query Match 72.4%; Score 21; DB 11; Length 215;  
Best Local Similarity 82.8%; Pred. No. 16;  
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 CGGTGAAGTTGACAGGATCCCAAGGTGAAG 29  
Db 201 CGGTGAAGATGACAGGATCCATGAGGAAG 173

RESULT 4  
US-10-225-567A-521/c  
; Sequence 521, Application US/10225567A  
; Publication No. US20030113798A1  
; GENERAL INFORMATION:

; APPLICANT: LifeSpan Biosciences  
; APPLICANT: Brown, Joseph P.  
; APPLICANT: Burmer, Glenn A.  
; APPLICANT: Roush, Christine L.  
; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPT.  
; FILE REFERENCE: 1920-4-4  
; CURRENT APPLICATION NUMBER: US/10/225,567A  
; CURRENT FILING DATE: 2001-12-19  
; PRIOR APPLICATION NUMBER: 60/257,144  
; PRIOR FILING DATE: 2000-12-19  
; NUMBER OF SEQ ID NOS: 2292  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 521  
; LENGTH: 5659  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-225-567A-521

Query Match 72.4%; Score 21; DB 14; Length 5659;  
Best Local Similarity 82.8%; Pred. No. 19;  
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 CGGTGAAGTTGACAGGATCCCAAGGTGAAG 29  
Db 2663 CGGTGAAGATGACAGGATCCATGAGGAAG 2635

RESULT 5  
US-09-844-653-32/c  
; Sequence 32, Application US/09844653  
; Publication No. US20030054347A1  
; GENERAL INFORMATION:  
; APPLICANT: Richards, Julia  
; APPLICANT: Rozsa, Frank  
; TITLE OF INVENTION: Detecting and Treating Eye Disease  
; FILE REFERENCE: UM-06105  
; CURRENT APPLICATION NUMBER: US/09/844,653  
; CURRENT FILING DATE: 2001-04-27  
; NUMBER OF SEQ ID NOS: 173  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 32  
; LENGTH: 48841  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc feature  
; LOCATION: (2067)..(2067)  
; OTHER INFORMATION: N at this position can be a, c, t, or g.  
US-09-844-653-32

Query Match 72.4%; Score 21; DB 11; Length 48841;  
Best Local Similarity 82.8%; Pred. No. 22;  
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 CGGTGAAGTTGACAGGATCCCAAGGTGAAG 29  
Db 36302 CGGTGAAGATGACAGGATCCATGAGGAAG 36274

RESULT 6  
US-09-987-190-13  
; Sequence 13, Application US/09987190  
; Patent No. US20020058293A1  
; GENERAL INFORMATION:  
; APPLICANT: TAKESAKO, Kazutoh  
; APPLICANT: MIZUTANI, Shigetoshi  
; APPLICANT: ENDO, Masahiro  
; APPLICANT: KATO, Ikunoshin  
; TITLE OF INVENTION: FUNGAL ANTIGENS AND PROCESS FOR PRODUCING THE SAME  
; FILE REFERENCE: 1422-0502P  
; CURRENT APPLICATION NUMBER: US/09/987,190  
; CURRENT FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: 09/262,856





; APPLICANT: Eroshkin, Alexey M  
; APPLICANT: Lemieux, Sebastien M  
; TITLE OF INVENTION: Identification of Essential Genes in *Aspergillus fumigatus* and  
; TITLE OF INVENTION: Methods of Use  
; FILE REFERENCE: 10182-018-999  
; CURRENT APPLICATION NUMBER: US/10/128,714  
; CURRENT FILING DATE: 2002-04-23  
; PRIOR APPLICATION NUMBER: US 60/285,697  
; PRIOR FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: US 60/287,066  
; PRIOR FILING DATE: 2001-04-27  
; PRIOR APPLICATION NUMBER: US 60/295,890  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US 60/303,899  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: US 60/316,362  
; PRIOR FILING DATE: 2001-08-31  
; NUMBER OF SEQ ID NOS: 8603  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5007  
; LENGTH: 2807  
; TYPE: DNA  
; ORGANISM: *Aspergillus fumigatus*  
US-10-128-714-5007

Query Match 64.8%; Score 18.8; DB 14; Length 2807;  
Best Local Similarity 90.9%; Pred. No. 1.6e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 7 AGTTGACAGGATCCAGGTGAA 28  
|||||  
Db 340 AGTTGACAGGATTCATGTTGAA 361

## RESULT 15

US-10-016-253-13/c  
; Sequence 13, Application US/10016253  
; Publication No. US20030124534A1  
; GENERAL INFORMATION:  
; APPLICANT: Rastelli, Luca  
; APPLICANT: Gould-Rothberg, Bonnie  
; APPLICANT: Murphey, Ryan  
; TITLE OF INVENTION: Method of Detecting and Treating Tuberous Sclerosis  
; TITLE OF INVENTION: Complex Associated Disorders  
; FILE REFERENCE: 21402-042  
; CURRENT APPLICATION NUMBER: US/10/016,253  
; CURRENT FILING DATE: 2002-08-28  
; PRIOR APPLICATION NUMBER: 60/254,268  
; PRIOR FILING DATE: 2000-12-08  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 713  
; TYPE: DNA  
; ORGANISM: *Homo sapiens*  
US-10-016-253-13

Query Match 64.1%; Score 18.6; DB 14; Length 713;  
Best Local Similarity 84.0%; Pred. No. 1.8e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 GTGAAGTTGACAGATCCAGGTGA 27  
|||||  
Db 132 GTGAAGTTACAGGAACAGGTGA 108

Search completed: October 27, 2003, 12:19:07  
Job time : 1448 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2003, 10:20:36 ; Search time 62 Seconds  
(without alignments)  
206.453 Million cell updates/sec

Title: US-10-043-418-5

Perfect score: 29

Sequence: 1 cggtaggttgacaggtccaaggtgaag 29

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTUS COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	25.8	89.0	2752	1	US-08-430-925A-3
C 2	21.4	72.4	5610	4	US-09-262-537-57
C 3	20.0	69.0	944	4	US-09-262-856A-13
C 4	20.0	69.0	1750	4	US-09-262-856A-7
C 5	18.4	63.4	1508	4	US-09-518-914-1
C 6	18.4	63.4	1541	4	US-09-518-914-3
C 7	18.0	62.1	51952	3	US-08-947-823-1
C 8	18.0	62.1	1830121	4	US-09-557-884-1
C 9	18.0	62.1	1830121	4	US-09-643-990A-1
C 10	17.8	61.4	867	4	US-09-107-532A-3466
C 11	17.8	61.4	1164	1	US-08-445-640-5
C 12	17.8	61.4	1164	3	US-08-170-558-5
C 13	17.8	61.4	1164	3	US-08-447-314-5
C 14	17.8	61.4	1164	3	US-08-445-461-5
C 15	17.8	61.4	2742	1	US-08-286-305A-2
C 16	17.8	61.4	2742	2	US-08-441-104A-2
C 17	17.8	61.4	2742	2	US-08-440-816A-2
C 18	17.8	61.4	2742	3	US-09-417-381A-2
C 19	17.8	61.4	3611	1	US-08-445-640-1
C 20	17.8	61.4	3611	3	US-08-170-558-1
C 21	17.8	61.4	3611	3	US-08-447-314-1
C 22	17.8	61.4	3611	3	US-08-445-461-1
C 23	17.8	61.4	5693	4	US-09-262-537-19
C 24	17.8	61.4	99500	4	US-09-798-096-10
C 25	17.4	60.0	1521	1	US-08-436-855A-3
C 26	17.4	60.0	1521	2	US-07-938-154-9
C 27	17.4	60.0	1521	5	PCT-US91-02311-9

C 28	17.4	60.0	1854	3	US-08-923-454A-13	Sequence 13, Appl
C 29	17.4	60.0	1902	4	US-09-620-312D-1057	Sequence 1057, Ap
C 30	17.4	60.0	2448	4	US-08-487-556-13	Sequence 13, Appl
C 31	17.4	60.0	2448	4	US-08-660-451A-13	Sequence 13, Appl
C 32	17.4	60.0	2450	2	US-08-466-589-9	Sequence 9, Appl
C 33	17.4	60.0	2450	2	US-08-700-636-9	Sequence 9, Appl
C 34	17.4	60.0	2450	3	US-08-467-574-9	Sequence 9, Appl
C 35	17.4	60.0	2450	4	US-09-217-345-9	Sequence 9, Appl
C 36	17.4	60.0	4403785	3	US-09-103-840A-2	Sequence 1, Appl
C 37	17.4	60.0	4411529	3	US-09-103-840A-1	Sequence 1, Appl
C 38	17.2	59.3	707	4	US-09-620-312D-709	Sequence 709, App
C 39	17.0	58.6	222	3	US-09-221-298-70	Sequence 70, Appl
C 40	17.0	58.6	291	4	US-09-702-705-365	Sequence 365, App
C 41	17.0	58.6	291	4	US-09-736-457-365	Sequence 365, App
C 42	17.0	58.6	350	3	US-08-888-077A-35	Sequence 35, Appl
C 43	17.0	58.6	493	1	US-08-592-126-117	Sequence 117, App
C 44	17.0	58.6	493	4	US-09-168-595-117	Sequence 117, App
C 45	17.0	58.6	597	3	US-09-385-982-277	Sequence 277, App

ALIGNMENTS

RESULT 1  
US-08-430-925A-3/c  
; Sequence 3, Application US/08430925A  
; Patent No. 5763252  
; GENERAL INFORMATION:  
; APPLICANT: Skadsen, Ronald W  
; APPLICANT: Tibbot, Brian K  
; TITLE OF INVENTION: Cloned Alpha-Glucosidase from Barley  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: 1 South Finckney Street  
; CITY: Madison  
; STATE: WI  
; COUNTRY: US  
; ZIP: 53703  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/430,925A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J  
; REGISTRATION NUMBER: 27,386  
; REFERENCE/DOCKET NUMBER: 960296.93171  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 608-251-5000  
; TELEFAX: 608-251-9166  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2752 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 42..2675  
; US-08-430-925A-3

Query Match 89.0%; Score 25.8; DB 1; Length 2752;  
Best Local Similarity 93.1%; Pred. No. 0, 032;  
Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CGTGAAGTTCACAGGATCCAAAGGTGAAG 29  
|||||

Db 1062 CGGTGAAGTTGACACGGTCCAAAGGTGAAG 1034

## RESULT 2

US-09-262-537-57/c  
; Sequence 57, Application US/09262537  
; Patent No. 6479256

## ; GENERAL INFORMATION:

; APPLICANT: Hayflick, Joel  
; TITLE OF INVENTION: Lectomedin Materials and Methods

; FILE REFERENCE: 27866/35307

; CURRENT APPLICATION NUMBER: US/09/262,537

; CURRENT FILING DATE: 1999-03-04

; EARLIER APPLICATION NUMBER: 60/076,782

; EARLIER FILING DATE: 1998-03-04

; NUMBER OF SEQ ID NOS: 64

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 57

; LENGTH: 5610

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (281)..(4687)

US-09-262-537-57

Query Match 72.4%; Score 21; DB 4; Length 5610;

Best Local Similarity 82.8%; Pred. No. 4.3;

Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGGTGAAGTTGACAGGATCCAAAGGTGAAG 29

||||| ||||| ||||| ||||| |||||

Db 2630 CGGTGAAGTGCACAGGATCCCAAGGTGAAG 2602

## RESULT 3

US-09-262-856A-13

; Sequence 13, Application US/09262856A

; Patent No. 6333164

## ; GENERAL INFORMATION:

; APPLICANT: TAKESAKO, Kazutoh

; APPLICANT: MIZUTANI, Shigetoshi

; APPLICANT: ENDO, Masahiro

; APPLICANT: KATO, Ikuroshin

; TITLE OF INVENTION: FUNGAL ANTIGENS AND PROCESS FOR PRODUCING THE SAME

; FILE REFERENCE: 1422-372P

; CURRENT APPLICATION NUMBER: US/09/262,856A

; CURRENT FILING DATE: 1999-03-04

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 13

; LENGTH: 944

; TYPE: DNA

; ORGANISM: Candida albicans

US-09-262-856A-13

Query Match 69.0%; Score 20; DB 4; Length 944;

Best Local Similarity 82.1%; Pred. No. 8.8;

Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CGGTGAAGTTGACAGGATCCAAAGGTGAAG 29

||||| ||||| ||||| ||||| |||||

Db 697 CGGTGAAGTGCACAGGATCCCAAGGTGAAG 724

## RESULT 4

US-09-262-856A-7

; Sequence 7, Application US/09262856A

; Patent No. 6333164

## ; GENERAL INFORMATION:

; APPLICANT: TAKESAKO, Kazutoh

; APPLICANT: MIZUTANI, Shigetoshi

; APPLICANT: ENDO, Masahiro

; APPLICANT: KATO, Ikunoshin  
; TITLE OF INVENTION: FUNGAL ANTIGENS AND PROCESS FOR PRODUCING THE SAME

; FILE REFERENCE: 1422-372P

; CURRENT APPLICATION NUMBER: US/09/262,856A

; CURRENT FILING DATE: 1999-03-04

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 7

; LENGTH: 1750

; TYPE: DNA

; ORGANISM: Candida albicans

US-09-262-856A-7

Query Match 69.0%; Score 20; DB 4; Length 1750;

Best Local Similarity 82.1%; Pred. No. 9.7;

Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 GGTGAAGTTGACAGGATCCCAAGGTGAAG 29

||||| ||||| ||||| ||||| |||||

Db 877 GGTGAAGTGCACAGGATCCCAAGGTGAAG 904

## RESULT 5

US-09-518-914-1/c

; Sequence 1, Application US/09518914

; Patent No. 6413731

## ; GENERAL INFORMATION:

; APPLICANT: Borowsky, Beth E.

; APPLICANT: Ogozalek, Kristine L.

; APPLICANT: Lakhani, Parul P.

; APPLICANT: Adham, Nika

; TITLE OF INVENTION: DNA ENCODING SNORF36a AND SNORF36b RECEPTORS

; FILE REFERENCE: 59138-A/JPW

; CURRENT APPLICATION NUMBER: US/09/518,914

; CURRENT FILING DATE: 2000-03-03

; EARLIER APPLICATION NUMBER: US 09/303,593

; EARLIER FILING DATE: 1999-05-03

; NUMBER OF SEQ ID NOS: 44

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 1508

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-518-914-1

Query Match 63.4%; Score 18.4; DB 4; Length 1508;

Best Local Similarity 78.8%; Pred. No. 47;

Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 GGTGAAGTTGACAGGATCCCAAGGTGAAG 29

||||| ||||| ||||| ||||| |||||

Db 396 GGTGAAGAGACAGGGGCGCTGGGTGAAG 369

## RESULT 6

US-09-518-914-3/c

; Sequence 3, Application US/09518914

; Patent No. 6413731

## ; GENERAL INFORMATION:

; APPLICANT: Borowsky, Beth E.

; APPLICANT: Ogozalek, Kristine L.

; APPLICANT: Lakhani, Parul P.

; APPLICANT: Adham, Nika

; TITLE OF INVENTION: DNA ENCODING SNORF36a AND SNORF36b RECEPTORS

; FILE REFERENCE: 59138-A/JPW

; CURRENT APPLICATION NUMBER: US/09/518,914

; CURRENT FILING DATE: 2000-03-03

; EARLIER APPLICATION NUMBER: US 09/303,593

; EARLIER FILING DATE: 1999-05-03

; NUMBER OF SEQ ID NOS: 44

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3

; LENGTH: 1541

TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-518-914-3

Query Match 63.4%; Score 18.4; DB 4; Length 1541;  
Best Local Similarity 78.6%; Pred. No. 47;  
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 GGTGAAGTTGACAGGATCCCAAGGTGAAG 29  
||||| ||||| || |||||  
Db 429 GGTGAAGACAGAGGGGCGCTGGGTGAAG 402

## RESULT 7

US-08-947-823-1/c  
; Sequence 1, Application US/08947823  
; Patent No. 6114605  
; GENERAL INFORMATION:  
; APPLICANT: Williamson, Valerie M.  
; APPLICANT: Kaloshian, Isgouhi  
; APPLICANT: Yaghoobi, Jafar  
; APPLICANT: Bodeau, John  
; APPLICANT: Milligan, Stephen  
; TITLE OF INVENTION: Procedures and Materials for Conferring  
; TITLE OF INVENTION: Pest Resistance in Plants  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3634

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/947,823  
FILING DATE: 09-OCT-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US97/18802  
FILING DATE: 09-OCT-1997

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/028,191  
FILING DATE: 10-OCT-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Bastian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 023070-070210US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 51952 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)

## US-08-947-823-1

Query Match 62.1%; Score 18; DB 3; Length 51952;  
Best Local Similarity 80.8%; Pred. No. 1.2e+02;  
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 TGAAGTTGACAGGATCCCAAGGTGAAG 29  
||||| ||||| || |||||  
Db 4207 TGTACTTGACAGTATAAAGGTGAAG 4182

## RESULT 8

US-09-557-884-1/c  
; Sequence 1, Application US/09557884  
; Patent No. 6506581  
; GENERAL INFORMATION:

APPLICANT: Fleischmann et al.  
TITLE OF INVENTION: The Nucleotide sequence of  
the Haemophilus influenzae Rd Genome, Fragments  
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: MD  
COUNTRY: USA  
ZIP: 20850

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3 1/2 inch diskette  
COMPUTER: Dell Pentium  
OPERATING SYSTEM: MS DOS v6.22  
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/557,884  
FILING DATE: 25-Apr-2000  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/476,102  
FILING DATE: JUN-5-1995

ATTORNEY/AGENT INFORMATION:  
NAME: Michelle S. Marks

REGISTRATION NUMBER: 41,971  
REFERENCE/DOCKET NUMBER: PB186P3

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 301-309-8504

TELEFAX: 301-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1830121 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

## US-09-557-884-1

Query Match 62.1%; Score 18; DB 4; Length 1830121;  
Best Local Similarity 80.8%; Pred. No. 1.4e+02;  
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 TGAAGTTGACAGGATCCCAAGGTGAAG 29  
||||| ||||| || |||||  
Db 16163 TGAAGTTGAATGACCAATGGTGAAG 16138

## RESULT 9

US-09-643-990A-1/c

; Sequence 1, Application US/09643990A  
; Patent No. 6528289

GENERAL INFORMATION:

APPLICANT: Robert D. Fleischmann  
Mark D. Adams  
Owen White  
Hamilton O. Smith  
J. Craig Venter

TITLE OF INVENTION: The Nucleotide sequence of  
the Haemophilus influenzae Rd Genome, Fragments  
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville,  
STATE: MD

```
;
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/445,461
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/170558
; FILING DATE: 20-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/157563
; FILING DATE: 23-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 854C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1164 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-445-461-5
```

```
Query Match 61.4%; Score 17.8; DB 3; Length 1164;
Best Local Similarity 75.9%; Pred. No. 82;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 CGGTGAAGTTGACAGGATCCAAAGTGGAAG 29
Db 26 CGGTGAAGCTGACAGTGTCTCAGGGGCAG 54
```

```
RESULT 15
US-08-286-305A-2
; Sequence 2, Application US/08286305A
; Patent No. 5766863
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Mark, Melanie R.
; APPLICANT: Sadick, Michael D.
; APPLICANT: Shelton, David L.
; APPLICANT: Wong, Wai Lee Tan
; TITLE OF INVENTION: KINASE RECEPTOR ACTIVATION ASSAY
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/286,305A
; FILING DATE: 05-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/170558
; FILING DATE: 20-DEC-1993
```

```
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/157563
; FILING DATE: 23-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER: 00,000
; REFERENCE/DOCKET NUMBER: 854C1P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1994
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2742 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-286-305A-2

Query Match 61.4%; Score 17.8; DB 1; Length 2742;
Best Local Similarity 75.9%; Pred. No. 94;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 CGGTGAAGTTGACAGGATCCAAAGTGGAAG 29
Db 146 CGGTGAAGCTGACAGTGTCTCAGGGGCAG 174

Search completed: October 27, 2003, 11:11:58
Job time : 80 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2003, 10:20:33 ; Search time 281 Seconds  
(without alignments)  
278.590 Million cell updates/sec

Title: US-10-043-418-5

Perfect score: 29

Sequence: 1 cggtagaattgacaggatccaagtggaag 29

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_19Jun03.\*  
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*  
25: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	25.8	89.0	2752	19	Barley alpha-glucosidase
C 2	21	72.4	525	21	Arabidopsis thaliana
C 3	21	72.4	3642	21	Human lectomedin-2
C 4	21	72.4	4091	21	Human lectomedin-2
C 5	21	72.4	4422	21	Human G protein-co
C 6	21	72.4	5610	20	Human lectomedin-2
C 7	21	72.4	5610	21	Originally determi
C 8	21	72.4	5610	21	Resequenced human

C 9	72.4	5659	21	AAZ87686	Human G protein-co
C 10	72.4	5659	25	ABZ42867	Human latrophilin-
C 11	69.0	764	22	AAH01294	Aspergillus nidula
C 12	69.0	944	19	AAV20827	Candida albicans f
C 13	69.0	1476	24	ABZ32400	Candida albicans f
C 14	69.0	1750	19	AAV20821	Candida albicans f
C 15	65.5	41	22	AAH73409	Human RS3 protein
C 16	65.5	777	24	ABQ66911	Human haemoglobin
C 17	65.5	2744	24	ABK86270	Common buckwheat a
C 18	65.5	2833	22	AAI93556	Human polynucleoti
C 19	65.5	2860	24	ABL51447	Buckwheat alpha-gl
C 20	65.5	3205	22	AAH73404	Human RS3 protein
C 21	65.5	3811	23	AAH91658	DNA encoding novel
C 22	65.5	4213	25	ABA00661	Human ENZM-9 cDNA,
C 23	65.5	4808	24	ABZ21063	4Fe-4S redox prote
C 24	64.8	2656	25	ABT17649	Aspergillus fumiga
C 25	64.8	2807	25	ABT19463	Aspergillus fumiga
C 26	64.1	713	24	ABS52774	Murine tuberosus sc
C 27	63.4	60	24	ABN58583	Human spliced tran
C 28	63.4	328	23	ABV15724	Human prostate exp
C 29	63.4	375	22	ABA17327	Human nervous syst
C 30	63.4	576	22	AAI43314	Probe #12000 used
C 31	63.4	616	23	ABV59122	Human prostate exp
C 32	63.4	823	22	AAH42812	Human G Protein-Co
C 33	63.4	1032	22	AAH01281	Exophiala moniliae
C 34	63.4	1046	22	AAH01286	Exophiala dermatit
C 35	63.4	1083	24	ABA96751	Acaryochloris mari
C 36	63.4	1085	22	AAH00466	Wangiella dermatit
C 37	63.4	1095	22	AAH00444	Exophiala jeanselm
C 38	63.4	1099	22	AAH00794	Phialophora verruc
C 39	63.4	1138	22	AAH00798	Trichophyton menta
C 40	63.4	1146	22	AAH00443	Cryptococcus albid
C 41	63.4	1149	22	AAH00464	Trichophyton menta
C 42	63.4	1155	22	AAH00797	Trichophyton tonsu
C 43	63.4	1244	22	AAH01280	Aspergillus versic
C 44	63.4	1343	22	AAH01290	Cryptococcus laure
C 45	63.4	1425	25	ABZ36367	Human secretory po

## ALIGNMENTS

RESULT 1  
AAV11736/c  
ID AAV11736 standard; cDNA; 2752 BP.

XX AC AAV11736;

DT 07-AUG-1998 (first entry)

XX Barley alpha-glucosidase cDNA.

XX Alpha-glucosidase; barley; starch grain hydrolysis; alpha-amylase;  
XX Glucoamylase; industry; germplasm; hydrolytic enzyme; ds.

OS Hordeum vulgare.

XX Key Location/Qualifiers

FT CDS 42..2675

FT /\*tag= a

FT /product= alpha-glucosidase

XX US5763252-A.

XX 09-JUN-1998.

XX 28-APR-1995; 95US-0430925.

XX 28-APR-1995; 95US-0430925.

XX (WISC ) WISCONSIN ALUMNI RES FOUND.

XX Skadsen RW, Tibbot BK;

```

XX WPI: 1998-347329/30.
DR P-PSDB; AAW59040.
XX
PT DNA encoding barley alpha-glucosidase protein - useful for producing
PT recombinant protein to increase rate of starch grain hydrolysis when
PT used with alpha amylase
XX
PS Claim 4; Col 15-22; 19pp; English.
XX
CC This cDNA sequence encodes a novel barley alpha-glucosidase protein.
CC Recombinant alpha-glucosidase can be used to increase the rate of starch
CC grain hydrolysis when used together with alpha-amylase or can supplement
CC glucoamylase in industrial starch hydrolysis systems. Useful DNA sequence
CC characteristics from this enzyme can be identified which can be used as
CC hybridisation probes for identifying germplasm with high levels of
CC efficient hydrolytic enzymes.
XX
SQ Sequence 2752 BP; 486 A; 974 C; 845 G; 447 T; 0 other;
Query Match      89.0%; Score 25.8; DB 19; Length 2752;
Best Local Similarity 93.1%; Pred. No. 0.17;
Matches 27; Conservative C; Mismatches 2; Indels 0; Gaps 0;
Qy      1 CGGTGAAGTTGACAGGATCCCAAGGTGAAG 29
        ||||| ||||| ||||| ||||| |||||
Db      1062 CGGTGAAGTTGACAGGATCCCAAGGTGAAG 1034

RESULT 2
AAC34879/c
ID AAC34879 standard; DNA; 525 BP.
XX
AC AAC34879;
XX
DT 17-OCT-2000 (first entry)
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 8222.
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; terminator; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0121180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.

```

```

PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.

```

PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145931.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.

PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 72.4%; Score 21; DB 21; Length 525;

Best Local Similarity 82.8%; Pred. No. 16;

Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CCGTGAAGTTGACAGGATCCAAAGTGAAG 29  
| | | | | | | | | | | | | | | | | | | |  
Db 106 CCGGGAATTTACAGGATCCAAAGTGAAG 78

#### RESULT 3

AAA96040/c  
ID AAA96040 standard; cDNA; 3642 BP.

XX AC AAA96040;

XX DT 26-JAN-2001 (first entry)

XX DE Human lectomedin-2 cDNA clone 2.4.

XX Human; lectomedin; seven transmembrane receptor protein;  
KW antiatherosclerotic; vasotrophic; cytostatic; gene therapy;  
KW atherosclerosis; restenosis; vascular disease; peripheral neuropathy;  
KW cancer; nerve regeneration; renal cystic epithelium;  
KW uterine implantation; ss.

XX OS Homo sapiens.

XX PN WO200052039-A2.

XX PD 08-SEP-2000.

XX PF 03-MAR-2000; 2000WO-US05934.

XX PR 04-MAR-1999; 99US-0262537.

XX PA (ICOS-) ICOS CORP.

XX PI Hayflick JS, Fox RD, Douangpanya J, Furi KD;

XX WPI: 2000-579274/54.  
P-PSDB; AAB15740.

PT Isolated human seven transmembrane receptor lectomedin polypeptide or  
its fragment, useful for identifying agents which can treat  
atherosclerosis, restenosis or vascular disease -

PS Example 1; Page 196-200; 241pp; English.

CC The present cDNA sequence was obtained and used in a procedure for the  
isolation of human seven transmembrane receptor lectomedin polypeptides.  
CC The polypeptides have a characteristic extracellular structure including  
CC lectin-binding, olfactomedin-like and mucin-like domains. Modulator  
CC compounds that inhibit the binding of lectomedin to a binding partner  
CC (galectin-3) may be used to treat conditions such as atherosclerosis,  
CC restenosis, vascular disease, peripheral neuropathy, cancer, nerve  
CC regeneration, and pathologies of the renal cystic epithelia and uterine  
CC implantation. Nucleotide sequences encoding the lectomedin polypeptides  
CC are useful in gene therapy. The lectomedin polynucleotides and  
CC polypeptides may be used to identify lectomedin binding partner  
CC compounds. The polynucleotides are useful for detecting lectomedin

CC expression in cells and for identifying genetic mutations in the  
 CC lectomedin genes. Antisense lectomedin nucleic acids are able to inhibit  
 CC expression of lectomedin genes.

SQ Sequence 3642 BP; 689 A; 1194 C; 1065 G; 694 T; 0 other;  
 Query Match 72.4%; Score 21; DB 21; Length 3642;  
 Best Local Similarity 82.8%; Pred. No. 22;  
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CGGTGAAGTTGACAGGATCCCAAGGTGAAG 29  
 ||||| ||||| ||||| ||||| |||||  
 Db 662 CGGTGAAGTGACAGGGTCCATGAGGAAG 634

RESULT 4  
 AAA96039/c  
 ID AAA96039 standard; cDNA; 4091 BP.  
 XX  
 AC AAA96039;  
 XX  
 DT 26-JAN-2001 (first entry)  
 XX  
 DE Human lectomedin-2 cDNA clone 2.1.  
 XX  
 KW Human; lectomedin; seven transmembrane receptor protein;  
 KW arteriosclerosis; vasotrophic; cytostatic; gene therapy;  
 KW atherosclerosis; restenosis; vascular disease; peripheral neuropathy;  
 KW cancer; nerve regeneration; renal cystic epithelium;  
 KW uterine implantation; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200052039-A2.  
 XX  
 PD 08-SEP-2000.  
 XX  
 PF 03-MAR-2000; 2000WO-US05934.  
 XX  
 PR 04-MAR-1999; 99US-0262537.  
 XX  
 PA (ICOS-) ICOS CORP.  
 XX  
 PI Hayflick JS, Fox RD, Douangpanya J, Puri KD;  
 XX  
 DR WPI: 2000-579274/54.  
 DR P-PSDB; AAB15739.

XX Isolated human seven transmembrane receptor lectomedin polypeptide or  
 PT its fragment, useful for identifying agents which can treat  
 PT atherosclerosis, restenosis or vascular disease -  
 XX  
 PS Example 1; Page 189-193; 241pp; English.  
 XX  
 CC The present cDNA sequence was obtained and used in a procedure for the  
 CC isolation of human seven transmembrane receptor lectomedin polypeptides.  
 CC The polypeptides have a characteristic extracellular structure including  
 CC lectin-binding, olfactomedin-like and mucin-like domains. Modulator  
 CC compounds that inhibit the binding of lectomedin to a binding partner  
 CC (galectin-3) may be used to treat conditions such as atherosclerosis,  
 CC restenosis, vascular disease, peripheral neuropathy, cancer, nerve  
 CC regeneration, and pathologies of the renal cystic epithelia and uterine  
 CC implantation. Nucleotide sequences encoding the lectomedin polypeptides  
 CC are useful in gene therapy. The lectomedin polynucleotides and  
 CC polypeptides may be used to identify lectomedin binding partner  
 CC compounds. The polynucleotides are useful for detecting lectomedin  
 CC expression in cells and for identifying genetic mutations in the  
 CC lectomedin genes. Antisense lectomedin nucleic acids are able to inhibit  
 CC expression of lectomedin genes.

SQ Sequence 4091 BP; 770 A; 1370 C; 1192 G; 759 T; 0 other;  
 Query Match 72.4%; Score 21; DB 21; Length 4091;  
 Best Local Similarity 82.8%; Pred. No. 22;  
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Best Local Similarity 82.8%; Pred. No. 23;  
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 Qy 1 CGGTGAAGTTGACAGGATCCCAAGGTGAAG 29  
 ||||| ||||| ||||| ||||| |||||  
 Db 2630 CGGTGAAGTGACAGGGTCCATGAGGAAG 2602

RESULT 5  
 AAZ87685/c  
 ID AAZ87685 standard; DNA; 4422 BP.  
 XX  
 AC AAZ87685;  
 XX  
 DT 15-MAY-2000 (first entry)  
 XX  
 DE Human G protein-conjugated receptor HK02631 encoding DNA.  
 XX  
 KW G protein; guanine nucleotide binding protein; human; brain;  
 KW G protein-conjugated receptor; gene therapy; HK02631; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200005264-A1.  
 XX  
 PD 03-FEB-2000.  
 XX  
 PF 22-JUL-1999; 99WO-JP03909.  
 XX  
 PR 23-JUL-1998; 98JP-0207579.  
 PR 07-AUG-1998; 98JP-0225060.  
 PR 06-OCT-1998; 98JP-0284328.  
 XX  
 PA (TAKE ) TAKEDA CHEM IND LTD.  
 PA (KAZU-) KAZUSA DNA RES INST.  
 XX  
 PI Ohara O, Nagase T, Nomura N, Hinuma S, Fujii R, Kitahara O;  
 PI Mogi S;  
 XX  
 DR WPI: 2000-182652/16.  
 DR P-PSDB; AAY77729.

XX New G protein conjugated receptor protein expressed in brain tissue for  
 PT screening potential agonists and antagonists of its binding to ligands  
 PT for use as drugs -  
 XX  
 PS Claim 4; Page 116-119; 123pp; Japanese.  
 XX  
 CC The invention provides G protein (guanine nucleotide binding protein)  
 CC -conjugated receptor proteins (AAY77727-29) expressed in brain tissue  
 CC and nucleic acids (AAZ87683-685) encoding the polypeptides. The  
 CC polypeptides and the methods are useful for identifying compounds which  
 CC are agonists or antagonists to the binding of the receptor to its ligand,  
 CC for use as drugs. DNA encoding all or part of the polypeptides is used  
 CC for the diagnosis of diseases and in gene therapy, and for the production  
 CC of transgenic animals for use as disease models. Antibodies recognizing  
 CC the receptor proteins or their fragments are also useful for disease  
 CC diagnosis. The present sequence represents a DNA encoding the G  
 CC protein-conjugated receptor protein HK02631.  
 XX  
 SQ Sequence 4422 BP; 855 A; 1517 C; 1267 G; 783 T; 0 other;

Query Match 72.4%; Score 21; DB 21; Length 4422;  
 Best Local Similarity 82.8%; Pred. No. 23;  
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 Qy 1 CGGTGAAGTTGACAGGATCCCAAGGTGAAG 29  
 ||||| ||||| ||||| ||||| |||||  
 Db 2365 CGGTGAAGTGACAGGGTCCATGAGGAAG 2337

RESULT 6  
 AAZ28008/c

ID AAZ28008 standard; DNA; 5610 BP.  
 XX AAZ28008;  
 AC  
 XX  
 DT 05-JAN-2000 (first entry)  
 XX  
 DE Human lectomedin-2 encoding DNA.  
 XX  
 XX Human; 7-transmembrane receptor; lectomedin; lectin-binding; mucin;  
 KW olfactomedin; cellular adhesion; atherosclerosis; gene therapy;  
 KW vascular disease; lectomedin-2; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9945111-A1.  
 XX  
 XX 10-SEP-1999.  
 PD  
 XX 04-MAR-1999; 99WO-US04676.  
 PF  
 XX 04-MAR-1998; 98US-0076782.  
 PR  
 XX (ICOS-) ICOS CORP.  
 PA  
 XX Hayflick JS;  
 PI  
 XX WPI; 1999-571596/48.  
 DR  
 XX P-PSDB; AAY41097.  
 DR  
 XX  
 XX New human lectomedin receptor polypeptide, used to identify specific  
 PT binding partners for treating e.g. vascular disease -  
 PT  
 XX Claim 1; Page 148-155; 166pp; English.  
 PS  
 XX The invention provides purified and isolated human 7-transmembrane  
 CC receptor lectomedin polypeptide or its fragments. The lectomedin  
 CC polypeptide comprises extracellular lectin-binding, olfactomedin-like  
 CC and mucin-like domains. The polypeptide can be produced by standard  
 CC recombinant methodology. The polypeptide is involved in cellular adhesion  
 CC and cytoplasmic metabolic pathways that are modulated by extracellular  
 CC signaling. Specific binding to lectomedin-1 expressed on smooth muscle  
 CC cells may be required for proliferation of these cells in  
 CC atherosclerosis. The polypeptide is used to raise specific antibodies,  
 CC and to identify specific binding agents that modulate (increase or  
 CC decrease) its activity. The lectomedin nucleic acids are used as source  
 CC of probes and primers, and of therapeutic antisense, ribozyme or triplex-  
 CC forming agents, and in gene therapy to restore deficient lectomedin  
 CC activity. Specific binding agents of lectomedin are used for treating  
 CC diseases that involve lectomedin activity, e.g. vascular diseases such as  
 CC atherosclerosis. The present sequence represents the DNA encoding human  
 CC lectomedin-2 polypeptide.  
 XX  
 XX Sequence 5610 BP; 1057 A; 1892 C; 1657 G; 1004 T; 0 other;  
 SQ  
 Query Match 72.4%; Score 21; DB 20; Length 5610;  
 Best Local Similarity 82.8%; Pred. No. 24;  
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 CGGTGAAGTTGACAGGATCCAGGTGAAG 29  
 |||||  
 Db 2630 CGGTGAAGATGACAGGTCATGAGGAAG 2602  
 RESULT 7  
 AAA96035/c  
 ID AAA96035 standard; cDNA; 5610 BP.  
 XX  
 AC AAA96035;  
 XX  
 DT 26-JAN-2001 (first entry)  
 XX  
 DE Originally determined sequence of human lectomedin-2 cDNA.  
 XX

KW Human; lectomedin; seven transmembrane receptor protein;  
 KW antiatherosclerotic; vasotropic; cytostatic; gene therapy;  
 KW atherosclerosis; restenosis; vascular disease; peripheral neuropathy;  
 KW cancer; nerve regeneration; renal cystic epithelium;  
 XX uterine implantation; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200052039-A2.  
 XX  
 PD 08-SEP-2000.  
 XX  
 XX 03-MAR-2000; 2000WO-US05934.  
 PF  
 XX 04-MAR-1999; 99US-0262537.  
 PR  
 XX (ICOS-) ICOS CORP.  
 PA  
 XX Hayflick JS, Fox RD, Douangpanya J, Puri KD;  
 PI  
 XX WPI; 2000-579274/54.  
 DR  
 XX P-PSDB; AAB15735.  
 DR  
 XX Isolated human seven transmembrane receptor lectomedin polypeptide or  
 PT its fragment, useful for identifying agents which can treat  
 PT atherosclerosis, restenosis or vascular disease -  
 XX  
 PS Example 1; Page 174-180; 241pp; English.  
 XX  
 CC The present cDNA sequence was obtained and used in a procedure for the  
 CC isolation of human seven transmembrane receptor lectomedin polypeptides.  
 CC The polypeptides have a characteristic extracellular structure including  
 CC lectin-binding, olfactomedin-like and mucin-like domains. Modulator  
 CC compounds that inhibit the binding of lectomedin to a binding partner  
 CC (galectin-3) may be used to treat conditions such as atherosclerosis,  
 CC restenosis, vascular disease, peripheral neuropathy, cancer, nerve  
 CC regeneration, and pathologies of the renal cystic epithelia and uterine  
 CC implantation. Nucleotide sequences encoding the lectomedin polypeptides  
 CC are useful in gene therapy. The lectomedin polynucleotides and  
 CC polypeptides may be used to identify lectomedin binding partner  
 CC compounds. The polynucleotides are useful for detecting lectomedin  
 CC expression in cells and for identifying genetic mutations in the  
 CC lectomedin genes. Antisense lectomedin nucleic acids are able to inhibit  
 CC expression of lectomedin genes.  
 XX  
 SQ Sequence 5610 BP; 1057 A; 1892 C; 1657 G; 1004 T; 0 other;  
 Query Match 72.4%; Score 21; DB 21; Length 5610;  
 Best Local Similarity 82.8%; Pred. No. 24;  
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 CGGTGAAGTTGACAGGATCCAGGTGAAG 29  
 |||||  
 Db 2630 CGGTGAAGATGACAGGTCATGAGGAAG 2602  
 RESULT 8  
 AAA96041/c  
 ID AAA96041 standard; cDNA; 5610 BP.  
 XX  
 AC AAA96041;  
 XX  
 DT 26-JAN-2001 (first entry)  
 XX  
 DE Resequenced human lectomedin-2 cDNA.  
 XX  
 KW Human; lectomedin; seven transmembrane receptor protein;  
 KW antiatherosclerotic; vasotropic; cytostatic; gene therapy;  
 KW atherosclerosis; restenosis; vascular disease; peripheral neuropathy;  
 KW cancer; nerve regeneration; renal cystic epithelium;  
 KW uterine implantation; ss.  
 XX  
 OS Homo sapiens.

XX WO200052039-A2.  
 PN  
 XX  
 PD  
 XX  
 XX 08-SEP-2000.  
 PF  
 XX 03-MAR-2000; 2000WO-US05934.  
 PR  
 XX 04-MAR-1999; 99US-0262537.  
 XX  
 XX (ICOS-) ICOS CORP.  
 PA  
 XX Hayflick JS, Fox RD, Douangpanya J, Puri KD;  
 PI P-PSDB; AAB15741.  
 DR  
 XX WPI: 2000-579274/54.  
 DR  
 XX  
 XX Isolated human seven transmembrane receptor lectomedin polypeptide or  
 PT its fragment, useful for identifying agents which can treat  
 PT atherosclerosis, restenosis or vascular disease -  
 XX  
 XX Claim 4; Page 203-210; 241pp; English.  
 PS  
 CC The present cDNA sequence was obtained and used in a procedure for the  
 CC isolation of human seven transmembrane receptor lectomedin polypeptides.  
 CC The polypeptides have a characteristic extracellular structure including  
 CC lectin-binding, olfactomedin-like and mucin-like domains. Modulator  
 CC compounds that inhibit the binding of lectomedin to a binding partner  
 CC (galactin-3) may be used to treat conditions such as atherosclerosis,  
 CC restenosis, vascular disease, peripheral neuropathy, cancer, nerve  
 CC regeneration, and pathologies of the renal cystic epithelia and uterine  
 CC implantation. Nucleotide sequences encoding the lectomedin polypeptides  
 CC are useful in gene therapy. The lectomedin polynucleotides and  
 CC polypeptides may be used to identify lectomedin binding partner  
 CC compounds. The polynucleotides are useful for detecting lectomedin  
 CC expression in cells and for identifying genetic mutations in the  
 CC lectomedin genes. Antisense lectomedin nucleic acids are able to inhibit  
 CC expression of lectomedin genes.  
 XX  
 SQ Sequence 5610 BP; 1057 A; 1893 C; 1656 G; 1004 T; 0 other;  
 Query Match 72.4%; Score 21; DB 21; Length 5610;  
 Best Local Similarity 82.8%; Pred. No. 24;  
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 CGGTGAAGTTGACAGGATCCAGGTGAAG 29  
 DB 2630 CGGTGAAGATGACAGGGTCCATGAGGAG 2602  
 RESULT 9  
 AA287686/c  
 ID AA287686 standard; DNA; 5659 BP.  
 XX  
 XX AA287686;  
 AC  
 XX 15-MAY-2000 (first entry)  
 DT  
 XX Human G protein-conjugated receptor related DNA sequence.  
 DE  
 XX G protein; guanine nucleotide binding protein; human; brain;  
 KW G protein-conjugated receptor; gene therapy; ds.  
 KW  
 XX Homo sapiens.  
 OS  
 XX WO200005264-A1.  
 PN  
 XX 03-FEB-2000.  
 PD  
 XX 22-JUL-1999; 99WO-JP03909.  
 PF  
 XX 23-JUL-1998; 98JP-0207579.  
 PR  
 XX 07-AUG-1998; 98JP-0225060.  
 PR  
 XX 06-OCT-1998; 98JP-0284328.

XX (TAKE ) TAKEDA CHEM IND LTD.  
 PA (KAZU-) KAZUSA DNA RES INST.  
 XX  
 PI Ohara O, Nagase T, Nomura N, Hinuma S, Fujii R, Kitahara O;  
 PI Mogi S;  
 XX  
 XX WPI: 2000-182652/16.  
 DR  
 XX New G protein conjugated receptor protein expressed in brain tissue for  
 PT screening potential agonists and antagonists of its binding to ligands  
 PT for use as drugs -  
 XX  
 XX Disclosure; Fig 21 to 24; 123pp; Japanese.  
 PS  
 CC The invention provides G protein (guanine nucleotide binding protein)  
 CC -conjugated receptor proteins (AA77727-29) expressed in brain tissue  
 CC and nucleic acids (AA287683-685) encoding the polypeptides. The  
 CC polypeptides and the methods are useful for identifying compounds which  
 CC are agonists or antagonists to the binding of the receptor to its ligand,  
 CC for use as drugs. DNA encoding all or part of the polypeptides is used  
 CC for the diagnosis of diseases and in gene therapy, and for the production  
 CC of transgenic animals for use as disease models. Antibodies recognizing  
 CC the receptor proteins or their fragments are also useful for disease  
 CC diagnosis. The present sequence represents a DNA related to the G  
 CC protein-conjugated receptor DNAs.  
 XX  
 SQ Sequence 5659 BP; 1077 A; 1896 C; 1664 G; 1022 T; 0 other;  
 Query Match 72.4%; Score 21; DB 21; Length 5659;  
 Best Local Similarity 82.8%; Pred. No. 24;  
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 CGGTGAAGTTGACAGGATCCAGGTGAAG 29  
 DB 2663 CGGTGAAGATGACAGGGTCCATGAGGAG 2635  
 RESULT 10  
 AB242867/c  
 ID AB242867 standard; DNA; 5659 BP.  
 XX  
 XX AB242867;  
 AC  
 XX 04-MAR-2003 (first entry)  
 DT  
 XX Human latrophilin-1 nucleotide SEQ ID NO:521.  
 DE  
 XX  
 KW G protein-coupled receptor; GPCR; antigenic peptide; gene therapy;  
 KW G protein-coupled receptor modulator; antibody; immune-related disease;  
 KW growth-related disease; cell regeneration-related disease; AIDS; cancer;  
 KW immunological-related cell proliferative disease; autoimmune disease;  
 KW Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;  
 KW osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;  
 KW graft versus host disease; Parkinson's disease; multiple sclerosis; pain;  
 KW psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;  
 KW mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;  
 KW hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;  
 KW ulcer; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200261087-A2.  
 PN  
 XX 08-AUG-2002.  
 PD  
 XX 19-DEC-2001; 2001WO-US50107.  
 PF  
 XX 19-DEC-2000; 2000US-257144P.  
 PR  
 XX (LIFE-) LIFESPAN BIOSCIENCES INC.  
 PA  
 XX Burner GC, Roush CL, Brown JP;  
 PI

XX WPI; 2003-046718/04.  
 DR P-PSDB; ABP82017.  
 XX  
 XX New isolated antigenic peptides e.g., for G protein-coupled receptors  
 PT (GPCR), useful for diagnosing and designing drugs for treating  
 PT conditions in which GPCRs are involved, e.g. AIDS, Alzheimer's disease,  
 PT cancer or autoimmune diseases -  
 XX  
 XX Disclosure; Fig 1; 523pp; English.  
 XX  
 CC The present invention describes antigenic peptides (I) comprising:  
 CC (a) any one of 1601 sequences (see ABP82019 to ABP83619) of 12-24 amino  
 CC acids. Also described: (1) an assay for the detection of a particular  
 CC G protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;  
 CC and (2) an isolated antibody having high specificity and high affinity  
 CC or avidity for a particular GPCR. (I) can be used as GPCR modulators and  
 CC in gene therapy. The antigenic peptides for GPCRs are useful in detecting  
 CC an antibody against a particular GPCR, and in the production of specific  
 CC antibodies. The peptides and antibodies are also useful for detecting the  
 CC presence or absence of corresponding GPCRs. The antigenic peptides for  
 CC GPCRs and antibodies are useful for diagnosing and designing drugs for  
 CC treating immune-related diseases, growth-related diseases, cell  
 CC regeneration-related diseases, immunological-related cell proliferative  
 CC diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,  
 CC atherosclerosis, bacterial, fungal, protozoan or viral infections,  
 CC osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute  
 CC inflammation, allergies, Crohn's disease, diabetes, graft versus host  
 CC disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,  
 CC anxiety, depression, schizophrenia, dementia, mental retardation, memory  
 CC loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension,  
 CC hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or  
 CC any other disorder in which GPCRs are involved. The antibodies may be  
 CC used in immunoassays and immunodiagnosis. ABZ42523 to ABZ42669 encode  
 CC GPCR proteins given in ABP81675 to ABP82018, which are used in the  
 CC exemplification of the present invention.  
 XX  
 SQ Sequence 5659 BP; 1077 A; 1896 C; 1664 G; 1022 T; 0 other;  
 Query Match 72.4%; Score 21; DB 25; Length 5659;  
 Best Local Similarity 82.8%; Pred. No. 24;  
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 CGGTGAAGTTGACAGGATCCCAAGGTGAAG 29  
 ||||| ||||| ||||| ||||| |||||  
 Db 2663 CGGTGAAGATGACAGGATCCCATGAGGAAG 2635  
 RESULT 11  
 AAH01294/c  
 ID AAHC1294 standard; DNA; 764 BP.  
 AC AAH01294;  
 XX  
 XX 24-JUL-2001 (first entry)  
 DT  
 XX Aspergillus nidulans nucleotide sequence SEQ ID NO:1285.  
 DE  
 XX Species specific; genus specific; family specific; probe; detection;  
 KW identification; algal; archaeal; bacterial; fungal; parasitical;  
 KW microorganism; diagnosis; translation elongation factor Tu; toxin;  
 KW translation elongation factor G; RecA recombinase; resistance;  
 KW catalytic subunit of proton-translocating ATPase; antimicrobial;  
 KW vaccine; primer; ds.  
 XX  
 XX Aspergillus nidulans.  
 OS  
 XX WO200123604-A2.  
 PN  
 XX 05-APR-2001.  
 PD  
 XX 28-SEP-2000; 2000WO-CA01150.  
 XX

PR 28-SEP-1999; 99CA-2283458.  
 PR 19-MAY-2000; 2000CA-2307010.  
 XX  
 XX (INFE-) INFECTION DIAGNOSTIC (IDI) INC.  
 XX  
 XX Bergeron MG, Boissinot M, Huletsky A, Menard C, Ouellette M;  
 PI Picard FJ, Roy PH;  
 XX WPI; 2001-245006/25.  
 XX  
 XX Nucleic acid sequences are used to generate universal probes and  
 PT primers which can be used to identify and detect the presence of algal,  
 PT archaeal, bacterial, fungal and parasitological species in a test sample -  
 XX  
 PS Claim 24; Page 1102-1103; 1580pp; English.  
 XX  
 CC The present invention describes a method for generating a repository of  
 CC nucleic acids of tuf, fus, atp and/or recA genes from which probes  
 CC and/or primers are derived. The method comprises amplifying the nucleic  
 CC acids of determined algal, archaeal, bacterial, fungal and parasitological  
 CC species with a combination of defined primer pairs. The method can be  
 CC used for producing probes and/or primers for detecting one or more  
 CC related microorganisms e.g. algae, archaea, bacteria, fungi and  
 CC parasites, for universal detection and for specific and ubiquitous  
 CC detection and identification of an algal, archaeal, bacterial, fungal  
 CC and parasitological species, genus, family and group. A nucleic acid (I)  
 CC obtained using the method of the invention can be used for the universal  
 CC detection of any bacterium, fungus or parasite in a sample and for the  
 CC detection of at least one antimicrobial agent resistance gene or at  
 CC least one toxin gene. hexA nucleic acids are used for the specific and  
 CC ubiquitous detection and for identification of Streptococcus pneumoniae.  
 CC (I) can be used to design a therapeutic agent which is effective against  
 CC microorganisms. Microbial species or genus or family or phylum or group  
 CC which can be detected include Abiotrophia adiacens, Bordetella sp.,  
 CC Corynebacterium sp., Enterobacteriaceae group, Escherichia coli,  
 CC Mycobacteriaceae family, Pseudomonads group, Streptococcus sp.,  
 CC Neisseria gonorrhoeae and Staphylococcus sp.. Using DNA based tests  
 CC provides faster results than substrate specificity tests as results can  
 CC be determined in an hour and improved accuracy is also achieved.  
 CC AAH00010 to AAH002304 represent nucleotide sequences and primers/probes  
 CC which are given in the exemplification of the present invention.  
 XX  
 SQ Sequence 764 BP; 159 A; 260 C; 184 G; 160 T; 1 other;  
 Query Match 69.0%; Score 20; DB 22; Length 764;  
 Best Local Similarity 82.1%; Pred. No. 48;  
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 2 GGTGAAGTTGACAGGATCCCAAGGTGAAG 29  
 ||||| ||||| ||||| ||||| |||||  
 Db 382 GGTGACGTTGGCAGGAGCGAAGGTGACG 355  
 RESULT 12  
 AAV20827  
 ID AAV20827 standard; cDNA to mRNA; 944 BP.  
 XX  
 XX AAV20827;  
 AC  
 XX 30-JUL-1998 (first entry)  
 DT  
 XX Candida albicans fungal antigen partial encoding cDNA SEQ ID NO:13.  
 DE  
 XX Candida albicans; vaccine; allergen; antigenic protein; fungal antigen;  
 KW immune response; infection; insoluble; ss.  
 KW  
 XX Candida albicans.  
 OS  
 XX WO9809990-A1.  
 PN  
 XX 12-MAR-1998.  
 PD  
 XX 29-AUG-1997; 97WO-JP03041.  
 XX

```

XX 31-MAR-1997; 97JP-0099775.
PR 04-SEP-1996; 96JP-0255400.
XX (TAKI ) TAKARA SHUZO CO LTD.
PA Endo M, Kato I, Mizutani S, Takesako K;
XX WPI; 1998-193553/17.
XX Fungal antigens comprising insoluble fraction of fungal cells -
PT useful for, e.g. stimulating immune response and treatment and
PT diagnosis of fungal infection(s)
XX Example 27; Page 84; 108pp; Japanese.
XX The present sequence represents a partial cDNA sequence from an example
CC of the present invention, which partially encodes an active vaccine
CC component or allergen derived from Candida albicans. The present
CC invention describes fungal antigens, comprising the insoluble fraction
CC of fungal cells having completely/partially removed cell walls. Also
CC described are nucleic acids encoding the antigens and a method for
CC producing the antigens. The antigens can be used for preparing
CC therapeutic compositions for stimulating immune response, e.g. as a
CC vaccine. They can be used for treatment of fungal infections, treatment
CC and prevention of allergies and diagnosis of fungal infections in
CC vertebrates. The vaccines are not live, and have low toxicity.
XX Sequence 944 BP; 304 A; 169 C; 220 G; 251 T; 0 other;
SQ Query Match 69.0%; Score 20; DB 19; Length 944;
Best Local Similarity 82.1%; Pred. No. 49;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 GGTGAAGTTGACAGATCCAGGTCGAAG 29
DB 697 GGTGAAGTTGTCACAGATCGAAGTGAAG 724

RESULT 13
ABZ32400
ID ABZ32400 standard; DNA; 1476 BP.
XX AC ABZ32400;
XX 30-JAN-2003 (first entry)
XX Candida albicans essential gene SEQ ID NO 6687.
XX Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
KW signal transduction; DNA replication; cell division; growth;
KW proliferation; Candida albicans; fungicide; antifungal; gene; ss.
XX Candida albicans.
OS WO200253728-A2.
PN 11-JUL-2002.
XX 26-DEC-2001; 2001WO-US49486.
XX 29-DEC-2000; 2000US-259128P.
PR 20-FEB-2001; 2001US-0792024.
PR 22-AUG-2001; 2001US-314050P.
XX (ELIT-) ELITRA PHARM INC.
PA Roemer T, Jiang B, Boone C, Bussey H, Ohlsen K;
PI WPI; 2002-566694/60.
XX P-PSDB; ABP73850.
XX Constructing strains for identifying gene products as effective targets
PT

PT for therapeutic intervention, by inactivating in the strain one allele
PT of a gene and placing other allele of the gene under conditional
PT expression
XX Claim 37; SEQ ID NO 6687; 167pp + Sequence Listing; English.
XX The invention relates to constructing (M1) a strain of diploid fungal
CC cells in which both alleles of a gene are modified, comprising modifying
CC one allele by insertion or replacement by a cassette having an
CC expressible selectable marker and modifying other allele by
CC recombination, of a promoter replacement fragment with a heterologous
CC promoter, (M1) that expression of the second allele is regulated by the
CC promoter. (M1) is useful for constructing a strain of diploid fungal
CC cells in which both alleles of a gene are modified. The diploid fungal
CC cells having both alleles modified are useful for identifying a gene that
CC is essential to the survival or growth of a fungus, a gene that
CC contributes to the virulence and/or pathogenicity of a fungus, a gene
CC that contributes to the resistance of a diploid fungus to an antifungal
CC agent, an antifungal agent that inhibits the growth of a diploid fungus
CC and for identifying a therapeutic agent for treatment of a mammalian
CC disease. (M1) is useful for identifying a compound which modulates the
CC activity of a gene product, preferably enzymatic activity, carbon
CC compound catabolism, biosynthetic, transporter, transcriptional,
CC translational, signal transduction, DNA replication and cell division
CC activity. The method is useful for identifying a compound having the
CC ability to inhibit growth or proliferation of C. albicans cells and for
CC treating infection by C. albicans. The present sequence is that of an
CC essential Candida albicans gene used in the method of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office.
XX Sequence 1476 BP; 470 A; 273 C; 329 G; 404 T; 0 other;
SQ Query Match 69.0%; Score 20; DB 24; Length 1476;
Best Local Similarity 82.1%; Pred. No. 53;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 GGTGAAGTTGACAGATCCAGGTCGAAG 29
DB 805 GGTGAAGTTGTCACAGATCGAAGTGAAG 832

RESULT 14
AAV20821
ID AAV20821 standard; cDNA to mRNA, 1750 BP.
XX AC AAV20821;
XX 30-JUL-1998 (first entry)
XX Candida albicans fungal antigen - allergen encoding cDNA SEQ ID NO:7.
XX Candida albicans; vaccine; allergen; antigenic protein; fungal antigen;
KW immune response; infection; insoluble; ds.
XX Candida albicans.
OS Key Location/Qualifiers
PH CDS 73..1548
FT /*tag= a
FT /product= "fungal antigen"
XX WO9809990-A1.
XX 12-MAR-1998.
XX 29-AUG-1997; 97WO-JP03041.
XX 31-MAR-1997; 97JP-0099775.
PR 04-SEP-1996; 96JP-0255400.
XX (TAKI ) TAKARA SHUZO CO LTD.
PA

```



XX Endo M, Kato I, Mizutani S, Takesako K;  
 XX WPI; 1998-193553/17.  
 DR P-PSDB; AAW53251.

XX Fungal antigens comprising insoluble fraction of fungal cells -  
 PT useful for, e.g. stimulating immune response and treatment and  
 PT diagnosis of fungal infection(s)

XX Claim 24; Page 80-81; 108pp; Japanese.

XX The present sequence encodes an active vaccine component or allergen  
 CC derived from *Candida albicans*, which is an antigenic protein. The  
 CC present invention describes fungal antigens, comprising the insolub-  
 CC fraction of fungal cells having completely/partially removed cell walls.  
 CC Also described are nucleic acids encoding the antigens and a method for  
 CC producing the antigens. The antigens can be used for preparing  
 CC therapeutic compositions for stimulating immune response, e.g. as a  
 CC vaccine. They can be used for treatment of fungal infections, treatment  
 CC and prevention of allergies and diagnosis of fungal infections in  
 CC vertebrates. The vaccines are not live, and have low toxicity.

XX Sequence 1750 BP; 578 A; 315 C; 354 G; 503 T; 0 other;

Query Match 69.0%; Score 20; DB 19; Length 1750;  
 Best Local Similarity 82.1%; Pred. No. 54;  
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 GGTGAGTTCACAGGATCCAAAGGTGAAG 29  
 ||||| ||||| ||||| ||||| |||||  
 Db 877 GGTGAGTTCACAGGATCCAAAGGTGAAG 904

## RESULT 15

AAH73409/c  
 ID AAH73409 standard; DNA; 41 BP.  
 XX AC AAH73409;  
 XX 25-SEP-2001 (first entry)  
 XX Human RS3 protein coding sequence probe #1.  
 DE Human; RS3 protein; cancer; haemopathy; immunological disease;  
 KW inflammation; HIV infection; gene therapy; probe; ss.  
 XX Homo sapiens.  
 OS WO200149736-A1.  
 PN 12-JUL-2001.  
 XX 25-DEC-2000; 2000WO-CN00662.  
 PF 29-DEC-1999; 99CN-C127231.  
 PR (JYFU-) UNIV FUDAN.  
 XX (SHAN-) SHANGHAI BIO DOOR GENE TECHNOLOGY LTD.  
 PA Mao Y, Xie Y;

WPI; 2001-441855/47.  
 XX Human RS3 protein and encoded polynucleotide, applicable in diagnosis  
 XX and treatment of malignant neoplasm, haemopathy, human immunodeficiency  
 PT virus infection, immunological diseases and inflammation -  
 XX Example 7; Page 20; 34pp; Chinese.

XX The present invention provides the protein and coding sequences of human  
 CC RS3 protein. The sequences can be used in the treatment of cancer,  
 CC haemopathy, HIV infection, immunological diseases and inflammation. The

CC present sequence is a probe for the coding sequence of the invention.  
 XX Sequence 41 BP; 10 A; 16 C; 7 G; 8 T; 0 other;

Query Match 65.5%; Score 19; DB 22; Length 41;  
 Best Local Similarity 81.5%; Pred. No. 83;  
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 GGTGAAGTTTGACAGGATCCAAAGGTGAA 28  
 ||||| ||||| ||||| ||||| |||||  
 Db 36 GGTGAAGTTTGACAGGATCCCAAGGTGAA 10

Search completed: October 27, 2003, 10:25:34  
 Job time : 284 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2003, 10:20:33 ; Search time 2677 Seconds  
(without alignments)  
443.175 Million cell updates/sec

Title: US-10-043-418-5

Perfect score: 29

Sequence: 1 cggtagagtgacaggatccaagtggaag 29

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*

30: em.htg.hum.\*

31: em.htg.inv.\*

32: em.htg.other.\*

33: em.htg.mus.\*

34: em.htg.pln.\*

35: em.htg.rod.\*

36: em.htg.mam.\*

37: em.htg.vrt.\*

38: em.sy.\*

39: em.htgo.hum.\*

40: em.htgo.mus.\*

41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	25.8	89.0	2738	8	HVU22450	U22450 Hordeum vul
C 2	25.8	89.0	2752	6	AR012240	AF118240 Sequence
C 3	25.8	89.0	7113	8	AF118226	AF118226 Hordeum v
C 4	22.8	78.6	119238	9	AC005608	AC005608 Homo sapi
C 5	22.8	78.6	148288	9	AC108121	AC108121 Homo sapi
C 6	22.8	78.6	163817	9	AC109448	AC109448 Homo sapi
C 7	21.6	74.5	165681	9	AC090646	AC090646 Homo sapi
C 8	21.6	74.5	166991	9	AC090884	AC090884 Homo sapi
C 9	21.6	74.5	199517	9	AL353741	AL353741 Human DNA
C 10	21.2	73.1	174728	2	AC120269	AC120269 Rattus no
C 11	21.2	73.1	196393	2	AC128316	AC128316 Rattus no
C 12	21.2	73.1	264340	2	AC112537	AC112537 Rattus no
C 13	21.2	73.1	271434	2	AC114059	AC114059 Rattus no
C 14	21	72.4	2709	8	BT002222	BT002222 Arabidops
C 15	21	72.4	2769	8	AY053414	AY053414 Arabidops
C 16	21	72.4	3482	9	BC007587	BC007587 Homo sapi
C 17	21	72.4	4283	8	AF014806	AF014806 Arabidops
C 18	21	72.4	4422	6	E41270	E41270 Novel G pro
C 19	21	72.4	5610	6	AR253254	AR253254 Sequence
C 20	21	72.4	5610	6	BD138014	BD138014 Lactomedi
C 21	21	72.4	5610	9	AF307079	AF307079 Homo sapi
C 22	21	72.4	5659	6	AX549236	AX549236 Sequence
C 23	21	72.4	5659	9	AB020628	AB020628 Homo sapi
C 24	21	72.4	33137	6	AX646721	AX646721 Sequence
C 25	21	72.4	33137	9	AB065919	AB065919 Homo sapi
C 26	21	72.4	93205	8	ATT22P22	AL162459 Arabidops
C 27	21	72.4	111338	8	ATF1612	AL162459 Arabidops
C 28	21	72.4	149144	9	AC069114	AC069114 Homo sapi
C 29	21	72.4	160733	8	OSJN00182	AL662981 Oryza sat
C 30	21	72.4	212842	9	AC102803	AC102803 Homo sapi
C 31	21	72.4	237931	9	AG022098	AC022098 Homo sapi
C 32	21	72.4	281000	3	TBCHRI2	AL929604 Trypanoso
C 33	21	72.4	308015	1	AE016783	AE016783 Pseudomon
C 34	20.6	71.0	206764	2	AC102605	AC102605 Mus muscu
C 35	20.6	71.0	219155	2	AC113540	AC113540 Mus muscu
C 36	20.6	71.0	22615	10	AC125044	AC125044 Mus muscu
C 37	20.6	71.0	228478	2	AC111421	AC111421 Rattus no
C 38	20.4	70.3	121195	2	CNS08C9M	AL772418 Oryza sat
C 39	20.4	70.3	132733	8	CNS08C44	AL772426 Oryza sat
C 40	20.4	70.3	194559	2	AC073668	AC073668 Mus muscu
C 41	20.4	70.3	198578	2	AC087138	AC087138 Mus muscu
C 42	20.2	69.7	9601	1	AE011893	AE011893 Xanthomon
C 43	20.2	69.7	110000	2	AC118787_3	Continuation (4 of
C 44	20.2	69.7	151645	2	AC137901_	Continuation (4 of
C 45	20.2	69.7	175773	2	AC113472	AC137901 Mus muscu
						AC113472 Mus muscu

# ALIGNMENTS

RESULT 1  
HVU22450/c  
LOCUS HVU22450 2738 bp mRNA linear PLN 27-JUN-1996  
DEFINITION Hordeum vulgare alpha-glucosidase mRNA, complete cds.  
ACCESSION U22450  
VERSION U22450.1 GI:944900  
KEYWORDS Hordeum vulgare subsp. vulgare  
SOURCE Hordeum vulgare subsp. vulgare  
ORGANISM Hordeum vulgare subsp. vulgare  
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Pooidae; Triticeae; Hordeum.  
REFERENCE 1 (bases 1 to 2738)  
AUTHORS Tibbot,B.K. and Skadsen,R.W.  
TITLE Molecular cloning and characterization of a gibberellin-inducible,

putative alpha-glucosidase gene from barley  
Plant Mol. Biol. 30 (2), 229-241 (1996)  
66178863  
PUBMED  
2 (bases 1 to 2738)  
REFERENCE  
Tibbot, B.K.  
Direct Submission  
Submitted (10-MAR-1995) Brian K. Tibbot, Agronomy, University of  
Wisconsin, 501 N. Walnut St., Madison, WI 53705, USA  
LOCATION/Qualifiers  
1. .2738  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="Morex"  
/db\_xref="taxon:112509"  
/chromosome="1"  
/map="near Amy2"  
/clone="pAGL.2737"  
/haplotype="2n"  
/tissue.type="aleurone"  
/dev\_stage="aleurones from deembryonated half-seeds from  
matured seeds"  
42. .2675  
/EC\_number="3.2.1.20"  
/note="gibberellin inducible"  
/codon\_start=1  
/product="alpha-glucosidase"  
/protein\_id="AA02985.1"  
/db\_xref="GI:944901"  
/translation="MATVGVLLCLCLCFAPRLCSSKEEGPLAARTVLAVAVTMEGA  
LRAEATGGRSTGDVQRLAVASLETDSRLVRITDADHPREVFPQDIIPRAPGQV  
LHDAFPASAPLQGRVLSGASDLVLTTHASPFRTVSRSTGTLFTAPLGFVRGK  
YLEVTSALPAGRASLYGEGHTKSSFLRHNDSPFLNADICASYVDVNLGSHFFVM  
DVAAGTAHGVLLSSNGMDVLYGGSVYTKVIGVLDFYFAGNPPLAVVQVQTLJ  
ARAPMYSFGFCRCYGLNVDLERVAVAKARIPLEVMNTDIDYMDGKDFTL  
DRVNTAEALRPFVDFMHPAAEFWAREISLFRPTIPVDGLWIDMNEISNFYPM  
INFVGNVPGDVFDFMHPAAEFWAREISLFRPTIPVDGLWIDMNEISNFYPM  
NALDDPYRINNDGTRINNKTVPLAVHVGTVTEYEHNLGLLEARATGRLVRD  
TGRRPFLVSRSTFVSGRTAYTGDNAATMGDLRYSINTMLSGFLGPMGMIADICG  
FNGTTELCGRWICLGAFYFSDHSAI FTVRELYLWPSVAAGKALGLRYQLLP  
YFYLWYEAHTGAPLAPLPSPYHDVATGVDRQFLGRLVSLVLEPGPTVDA  
YFPAGRWALDYSLAVATRGKHLVPADATVNVHLTGTLPLQOSALTTRARR  
TAFHLVALAEDGTASGYLFDGDSPEYGRSSDMSVRFNYKIPNKGAIKVKSEV  
HNSYASQRTLVISKVLMGHRSPAPKKLTVHNSAEVEASSAGTRYCNAGGLGVA  
HIGLSLVVGEFFELKVMYS"  
1335. .1361  
/note="encodes catalytic region"  
2676. .2738  
2714. .2719  
2738  
/note="14 A nucleotides"  
BASE COUNT 472 a 974 c 845 g 447 t  
ORIGIN  
misc\_feature  
3'UTR  
polyA\_signal  
polyA\_site  
Query Match 89.0%; Score 25.8; DB 8; Length 2738;  
Best Local Similarity 93.1%; Pred. No. 0.53;  
Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CGGTGAAGTTGACAGGATCCAGGTTGAAG 29  
|||||  
DB 1062 CGGTGAAGTTGACAGGTTGAAGTTGAAG 1034  
RESULT 2  
AR012240/c  
LOCUS  
DEFINITION  
Accession 3 from patent US 5763252.  
AR012240  
VERSION  
AR012240.1 GI:3970230  
KEYWORDS  
Unknown.  
SOURCE  
Unknown.  
ORGANISM  
Unclassified.

REFERENCE 1 (bases 1 to 2752)  
Skadsen, R.W. and Tibbot, B.K.  
Cloned alpha-glucosidase from barley  
Patent: US 5763252-A 3 09-JUN-1998;  
JOURNAL  
FEATURES  
Location/Qualifiers  
1. .2752  
/organism="unknown"  
BASE COUNT 486 a 974 c 845 g 447 t  
ORIGIN  
Query Match 89.0%; Score 25.8; DB 6; Length 2752;  
Best Local Similarity 93.1%; Pred. No. 0.53;  
Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CGGTGAAGTTGACAGGATCCAGGTTGAAG 29  
|||||  
DB 1062 CGGTGAAGTTGACAGGTTGAAGTTGAAG 1034  
RESULT 3  
AF118226/c  
LOCUS  
DEFINITION  
Hordeum vulgare high pI alpha-glucosidase (AGL97) gene, complete  
cds.  
ACCESSION AF118226  
VERSION AF118226.1 GI:8547061  
KEYWORDS  
SOURCE  
ORGANISM  
Hordeum vulgare subsp. vulgare  
Hordeum vulgare subsp. vulgare  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Hordeum.  
REFERENCE 1 (bases 1 to 7113)  
Frandsen, T.P., Lok, F., Mirgorodskaya, E., Roepstorff, P. and  
Stensson, B.  
Purification, enzymatic characterization, and nucleotide sequence  
of a high-isoelectric-point alpha-glucosidase from barley malt  
Plant Physiol. 123 (1), 275-286 (2000)  
MEDLINE  
20267959  
PUBMED  
10806244  
REFERENCE 2 (bases 1 to 7113)  
Lok, F.  
Direct Submission  
Submitted (06-JAN-1999) Carlsberg Laboratory, Carlsberg Research  
Center, Gl. Carlsbergvej 10, Valby 2500, Denmark  
JOURNAL  
FEATURES  
Location/Qualifiers  
1. .7113  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="genomic DNA"  
/cultivar="Igri"  
/db\_xref="taxon:112509"  
1. .7113  
/gene="AGL97"  
/notes="HVAGL97"  
join(1. .218, 4505. .5166, 5268. .5498, 5585. .7113)  
/gene="AGL97"  
/codon\_start=1  
/product="high pI alpha-glucosidase"  
/protein\_id="AAF76254.1"  
/db\_xref="GI:8547062"  
/translation="MATRSLLLCLCLCFAPRLCSSKEEGPLAAGGYRVSRAVDDG  
GRRLRAEAAATGASSTGDVQRLAVASLETDSRLVRITDADHPREVFPQDIIPRP  
APADVLHDAFPASAPLQGRVLSGASDLVLTTHASPFRTVSRSTGTLFTAPG  
LVPRDKYLEVTSALPAGRASLYGEGHTKSSFLRHNDSPFLNADICASYVDVNLG  
SHFFYMDVPPGCTAGHGVLLSSNGMDVLYGGSVYTKVIGVLDFYFAGNPPLAVVD  
QTILGRPAPMPYFSGHCRCYGLNVDLERVAVAKARIPLEVMNTDIDYMDGKDFTL  
FKDFTLDRVNTAEALRPFVDFMHPAAEFWAREISLFRPTIPVDGLWIDMNEISNFY  
EPNALDDPYRINNDGTRINNKTVPLAVHVGTVTEYEHNLGLLEARATGRLVRD  
LMDTGRPFLVSRSTFVSGRTAYTGDNAATMGDLRYSINTMLSGFLGPMGMIADICG  
ICGFNGTTELCGRWICLGAFYFSDHSAI FTVRELYLWPSVAAGKALGLRYQL  
LILPYFTLWYEAHTGAPLAPLPSPYHDVATGVDRQFLGRLVSLVLEPGPTVDA  
YFPAGRWALDYSLAVATRGKHLVPADATVNVHLTGTLPLQOSALTTRARR  
VDAYFPAGRWALDYSLAVATRGKHLVPADATVNVHLTGTLPLQOSALTTRARR

```

ARTAFHLLVALAEDGTASDGLFDLDCSEPMGCRSDWSMVRFSCEMGSDDGAIVKVKSE
VHNSVAQKRTLVISKVLMGHRSPAPKKLTVHNSAEVASSSAGTRYQNAGGIGG
VAHIGGLSLVVGREFELKAMSY"

BASE COUNT      1709 a   1935 c   1716 g   1753 t
ORIGIN

Query Match      89.0%; Score 25.8; DB 8; Length 7113;
Best Local Similarity 93.1%; Pred. No. 0.58;
Matches 27; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CGGTGAAGTTGACAGGATCCCAAGGTGAAG 29
      |||||
Db      5426 CGGTGAAGTTGACACGGTCCCAAGGTGAAG 5398

RESULT 4
AC005608/c
LOCUS      AC005608      119298 bp      DNA      linear      PRI 04-SEP-1998
DEFINITION Homo sapiens chromosome 5, PAC clone 53L17 (LBNL H152), complete
sequence.
ACCESSION      AC005608
VERSION      ACC05608.1 GI:3540170
KEYWORDS      HTG.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 119298)
AUTHORS      Church, D.M., Yang, J., Socian, M., Shiang, R. and Wasmuth, J.J.
TITLE      A high-resolution physical and transcript map of the Cri du chat
region of human chromosome 5p
JOURNAL      Genome Res. 7 (8), 787-801 (1997)
MEDLINE      97413160
PUBMED      9267803
REFERENCE      2 (bases 1 to 119298)
AUTHORS      Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,
Kadner, K., Miguel, T., Miller, C., Pitluck, S., Pollard, M.,
Rojeski, H., Subramanian, S. and Martin, C.H.
TITLE      Sequencing of human chromosome 5
JOURNAL      Unpublished
REFERENCE      3 (bases 1 to 119298)
AUTHORS      Rickes, D.O.
TITLE      Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) System
JOURNAL      Unpublished
AUTHORS      Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,
Kadner, K., Miguel, T., Miller, C., Pitluck, S., Pollard, M.,
Rojeski, H., Subramanian, S. and Martin, C.H.
TITLE      Direct Submission
JOURNAL      Submitted (04-SEP-1998) Human Genome Center, DOE Joint Genome
Institute, Lawrence Berkeley National Laboratory, MS 74-157,
Berkeley, CA 94720, U.S.A.
COMMENT      Sequence submitted by:
DOE Joint Genome Institute.
FEATURES
    source
        1..119298
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="5"
            /map="5q"
            /clone="PAC 53L17"
            /note="LBNL H152"
        991..1012
            /note="(T)22"
        /rpt_type=tandem
        /rpt_unit=A
    repeat_region
        complement (995..1179)
        /rpt_family="Alu"
    repeat_region
        complement (1238..1478)
        /rpt_family="Alu"
    repeat_region
        complement (3996..4092)
        /rpt_family="Alu"
    repeat_region
        complement (56348..57894)
        /rpt_family="MSTAR"
    repeat_region
        complement (57923..58284)
        /rpt_family="THE1"
    repeat_region
        complement (59006
        62501..62774
        /rpt_family="Alu"

```

```

/rpt_family="MLT1"
complement (4807..4926)
/rpt_family="MER42"
complement (7052..7666)
/rpt_family="L1"
8389..8676
/rpt_family="Alu"
8659..8691
/note="(A)33"
/rpt_type=tandem
/rpt_unit=A
complement (9668..9962)
/rpt_family="Alu"
11316..11369
/rpt_family="MLT1"
16149..16433
/rpt_family="Alu"
complement (20408..20538)
/note="GAIL 2 excellent exon, frame 2"
21280..24869
/rpt_family="L1"
25039..25369
/rpt_family="L1"
complement (30190..30464)
34356..34401
/note="(AC)23"
/rpt_type=tandem
/rpt_unit=AC
complement (35763..35869)
/note="GAIL 2 excellent exon, frame 2"
complement (38408..38595)
/rpt_family="MER7"
42337..42605
/rpt_family="Alu"
42852..42935
/note="GAIL 2 excellent exon, frame 2"
45174..45245
/rpt_family="Alu"
45358..45379
/note="(T)22"
/rpt_type=tandem
/rpt_unit=T
45690..45749
/note="GAIL 2 excellent exon, frame 0"
complement (47409..47554)
/rpt_family="MER33"
48205..48559
/rpt_family="THE1"
complement (49914..50183)
/rpt_family="Alu"
complement (50616..52070)
/rpt_family="MER7"
50905..51188
/rpt_family="Alu"
53050..53359
/rpt_family="Alu"
55278..55455
/rpt_family="Alu"
55782..56066
/rpt_family="Alu"
56049..56070
/note="(A)22"
/rpt_type=tandem
/rpt_unit=A
complement (56348..57894)
/rpt_family="MSTAR"
complement (57923..58284)
/rpt_family="THE1"
58788..59006
/rpt_family="Alu"
62501..62774
/rpt_family="Alu"

```





## ORIGIN

Query Match 74.5%; Score 21.6; DB 9; Length 165681;  
 Best Local Similarity 85.7%; Pred. No. 71;  
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 GGTGAAGTTGACAGGATCCAGGTGAAG 29  
 |||||

Db 92763 GGTGAAGTTGACAGGATCCAGGAAAAG 92736  
 |||||

## RESULT 8

AC090884/c  
 LOCUS AC090884 Homo sapiens chromosome 3 clone RP11-165B2 map 3p, complete  
 DEFINITION

AC090884 AC011607

VERSION AC090884.1 GI:13357222

## KEYWORDS

SOURCE HTG.

## ORGANISM

Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 166991)  
 Song L., Bao J., Bao Q., Bao W., Bian X., Cao T., Chen C., Chen J.,  
 Ding H., Dong W., Fan H., Feng X., Gong J., Guan Q., Gu X., Guo D.,  
 Guo Z., He L., Hu S., Huang F., Jin Y., Kang N., Li C., Li C.,  
 Li F., Li G., Li J., Li L., Li S., Li T., Liu Y., Liu N., Liu B.,  
 Liu Y., Li W., Li W., Li Y., Luo C., Luo J., Niu Y., Qi Q., Qi X.,  
 Song S., Sun M., Sun W., Sun Y., Tan X., Tao R., Wang H., Wang J.,  
 Wang J., Wang J., Wang L., Wang L., Wang R., Wang X., Wang X.,  
 Wang Y., Wu D., Wu Q., Xie F., Xuan Z., Xue Y., Yan C., Yang X.,  
 Yu B., Zeng Y., Zhang G., Zhang H., Zhang H., Zhang L., Zhang M.,  
 Zhang X., Zhang X., Zhang Y., Zhang Y., Zhang Z., Zhu S., Zhu N.,  
 Yu J. and Yang H.

## AUTHORS

Chromosome 3p genomic sequence  
 Unpublished  
 2 (bases 1 to 166991)  
 Song L., Bao J., Bao Q., Bao W., Bian X., Cao T., Chen C., Chen J.,  
 Ding H., Dong W., Fan H., Feng X., Gong J., Guan Q., Gu X., Guo D.,  
 Guo Z., He L., Hu S., Huang F., Jin Y., Kang N., Li C., Li C.,  
 Li F., Li G., Li J., Li L., Li S., Li T., Liu Y., Liu N., Liu B.,  
 Liu Y., Li W., Li W., Li Y., Luo C., Luo J., Niu Y., Qi Q., Qi X.,  
 Song S., Sun M., Sun W., Sun Y., Tan X., Tao R., Wang H., Wang J.,  
 Wang J., Wang J., Wang L., Wang L., Wang R., Wang X., Wang X.,  
 Wang Y., Wu D., Wu Q., Xie F., Xuan Z., Xue Y., Yan C., Yang X.,  
 Yu B., Zeng Y., Zhang G., Zhang H., Zhang H., Zhang L., Zhang M.,  
 Zhang X., Zhang X., Zhang Y., Zhang Y., Zhang Z., Zhu S., Zhu N.,  
 Yu J. and Yang H.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

Direct Submission  
 Submitted (16-MAR-2001) Human Genomic Center, Institute of  
 Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing  
 100101, P.R.China  
 On Mar 16, 2001 this sequence version replaced gi:9966194.  
 -----Genome Center  
 Center:Beijing Center  
 Center code:Beijing  
 Website:http://hgc.igtp.ac.cn  
 http://www.genomics.org.cn  
 Contact:hgc@igtp.ac.cn  
 -----Project Information  
 Center project name:11 project  
 Center clone name: RP11-165B2

## T-TITLE

## JOURNAL

## COMMENT

Sequencing vector: pUC18; 100% of reads  
 Chemistry: Dye-terminator; ET 5% of reads  
 Chemistry: Dye-terminator Big Dye; 45% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 173 bases at least Q40  
 Consensus quality: 419 bases at least Q30  
 Consensus quality: 482 bases at least Q20  
 Insert size: 490; sum-of-contigs  
 Quality coverage: 1.44x in Q20 bases,sum-of-contigs

## FEATURES

## source

Location/Qualifiers  
 1..166991  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="3"  
 /map="3p"  
 /clone="RP11-165B2"  
 42168 a 41024 c 39989 g 43810 t

## BASE COUNT

## ORIGIN

Query Match 74.5%; Score 21.6; DB 9; Length 166991;  
 Best Local Similarity 85.7%; Pred. No. 71;  
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 GGTGAAGTTGACAGGATCCAGGTGAAG 29  
 |||||

Db 124678 GGTGAAGTTGACAGGATCCAGGAAAAG 124651  
 |||||

## RESULT 9

AL353741/c

LOCUS AL353741

DEFINITION Human DNA sequence from clone Rp11-575C20 on chromosome 9, complete

sequence.

AL353741

VERSION AL353741.16 GI:12584694

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Submitted (28-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 requests: clonerequest@sanger.ac.uk  
 On Jan 28, 2001 this sequence version replaced gi:12580980.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.

This sequence has been finished according to sequence map criteria  
 as follows. An attempt is made to resolve all sequencing problems,  
 such as compressions and repeats, but not necessarily within known  
 annotated repeat sequence elements. Where the sequence is  
 ambiguous, there is an annotation using the 'unsure' feature key.  
 The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information  
 on the WORMPEP database can be found at

http://www.sanger.ac.uk/projects/C.elegans/wormpep This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping  
 Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr9  
 RP11-575C20 is from the library RPL1-11.2 constructed by the group  
 of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pBACe3.6

This sequence is the entire insert of clone Rp11-575C20 The true  
 left end of clone Rp11-279E1 is at 116738 in this sequence. The  
 true right end of clone Rp11-218I7 is at 84470 in this sequence.

## FEATURES

## source

Location/Qualifiers  
 1..199517  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="9"

[illegible]



```

misc_feature 6907. 7986
/notes="clone boundary
clone end:Sp6
site:MboI
end sequence:RXBQG14TV"
complement(169403..170490)
/notes="clone boundary
clone end:T7
end sequence:RXBQG14TV"
site:MboI
end sequence:RXBQG14TV"
/notes="wgs end_extension
clone end:T7"
BASE COUNT 44465 a 28574 c 29392 g 44439 t 27858 others
ORIGIN

Query Match 73.1%; Score 21.2; DB 2; Length 174728;
Best Local Similarity 88.5%; Pred. No. 1.1e+02;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GGTGAAGTTCACAGGATCAAGTGA 27
|||||
Db 31397 GGTGAAGTTCACAGGATCAAGTGA 31422
|||||

RESULT 11
AC128316 196393 bp DNA linear HTG 19-NOV-2002
LOCUS Rattus norvegicus clone CH230-333G1, WORKING DRAFT SEQUENCE, 2
DEFINITION
AC128316
VERSION AC128316.3 GI:25085897
KEYWORDS HTG: HTGS PHASE1; HTGS DRAFT: HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 196393)
Muzny, D.Marie., Metzker, M.Lee., Abrarzon, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabis, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebrgeorgis, F., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunnarsson, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensuwa, L., Louleed, H., Lozado, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Margum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montenayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Narkervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,

```

Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,  
Sanders, M., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smajs, D.,  
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
Valas, R., Vera, V., Villanana, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, R., Warren, R., Wei, X., White, F.,  
Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
Niederhauser, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
Weinstock, G. and Gibbs, R.A.  
Direct Submission  
Unpublished  
2 (bases 1 to 196393)  
Worley, K.C.  
Direct Submission  
Submitted (19-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 196393)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Nov 19, 2002 this sequence version replaced gi:23907853.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GRDP  
Center clone name: CH230-333G1  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 154698 bases at least Q40  
Consensus quality: 155755 bases at least Q30  
Consensus quality: 156631 bases at least Q20  
Estimated insert size: 157112; sum-of-contigs estimation  
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 195249: contig of 195249 bp in length  
\* 195250 195349: gap of unknown length  
\* 195350 196393: contig of 1044 bp in length.  
FEATURES  
Location/Qualifiers  
1..196393  
source

```

/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-333G1"
complement(1614..2468)
/note="clone boundary
clone_end:Sp6
site:
end_sequence:B2235011"
complement(176567..177394)
/note="clone boundary
clone_end:T7
site:
end_sequence:B2234990"
182181..183221
/note="wgs_end_extension
clone_end:T7"
194172..195249
/note="wgs_end_extension
clone_end:T7"
BASE COUNT 40837 a 37304 c 38177 g 41134 t 38941 others
ORIGIN

```

Query Match 73.1% Score 21.2; DB 2; Length 196393;  
 Best Local Similarity 88.5%; Pred. No. 1.1e+02;  
 Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 4 TGAAGTTCACAGGATCCCAAGGTGAAG 29
||||| ||||| ||||| ||||| |||||
Db 121029 TGAAGTGGAGGATCCCAAGGTGAAG 121054

```

```

RESULT i2
AC112537/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-40Pl3, *** SEQUENCING IN PROGRESS
***, 2 unordered pieces.
AC112537
AC112537.4 GI:30578442
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
SOURCE
Rattus norvegicus

```

```

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 264340)
Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alibrooks,S., Amin,A., Anguiano,D.,
Anvilebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswas,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Faves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Georgopoulos,E., Geer,K., Gill,R., Grady,M., Guerra,M., Guevara,W.,
Guanararte,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kovis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,B., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenshewa,L., Loulseghe,H., Lozada,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,

```

REFERENCE  
 AUTHORS

```

Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GJYM
Center clone name: CH230-40Pl3
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 231048 bases at least Q40
Consensus quality: 233413 bases at least Q30
Consensus quality: 235006 bases at least Q20
Estimated insert size: 252518; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

```

Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,B.,  
 Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,  
 Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,  
 Narkervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,  
 Nwankwelenh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,  
 Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,  
 Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,  
 Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,  
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,  
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,  
 Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,  
 Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,  
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,  
 Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,  
 Taylor,I., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmari,K.,  
 Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,  
 Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,  
 Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,  
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,  
 Yu,F., Zhang,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von  
 Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
 Weinstein,G. and Gibbs,R.A.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 264340)  
 Worley,K.C.  
 Direct Submission  
 Submitted (22-FEB-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 264340)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (13-MAY-2003) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On May 13, 2003 this sequence version replaced gi:23269061.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

COMMENT

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

\* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 260324: contig of 260324 bp in length  
 \* 260325 260424: gap of unknown length  
 \* 260425 264340: contig of 3916 bp in length.

## FEATURES

source  
 1. .264340  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-40P13"  
 misc\_feature  
 1. 1793  
 /note="wgs contig"  
 misc\_feature  
 255872. .257215  
 /note="wgs contig"  
 misc\_feature  
 258530. .260324  
 /note="wgs contig"

BASE COUNT 75767 a 46794 c 45644 g 67385 t 28750 others  
 ORIGIN

Query Match 73.1%; Score 21.2; DB 2; Length 264340;  
 Best Local Similarity 88.5%; Pred. No. 1.1e+02;  
 Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GGTGAAGTTCACGAGTCAAGGTCA 27  
 |||||  
 Db 88922 GGTGAAGTTCACGAGTCAAGGTCA 88897

## RESULT 13

AC114059/c

## LOCUS

DEFINITION Rattus norvegicus clone CH230-197C21, \*\*\* SEQUENCING IN PROGRESS

AC114059 271434 bp DNA linear HTG 13-MAY-2003

\*\*\*, 2 unordered pieces.

AC114059.5 GI:30581570

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

1 (bases 1 to 271434)

## REFERENCE

## AUTHORS

Muzny, D., Marie, M., Metzker, M., Lee, A., Adams, C., Alder, J.,  
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
 Biswal, K., Blais, J., Blankenburg, K., Blyth, P., Brown, M.,  
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
 Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,  
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
 Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
 Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,  
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
 Kowals, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
 Loresuhewa, L., Loubege, H., Lozano, R.J., Lu, X., Ma, J.,  
 Maheshwari, M., Mahindarne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
 Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,  
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,

Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
 Nwackelmele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,  
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
 Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,  
 Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,  
 Reilly, B., Reilly, M., Reuter, M., Richards, S., Riggs, F.,  
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,  
 Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsmar, S., Shen, H.,  
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,  
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
 Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,  
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
 Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K.,  
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
 Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
 Weinstock, G. and Gibbs, R.A.

Direct Submission

Unpublished

2 (bases 1 to 271434)

Worley, K.C.

Direct Submission

## JOURNAL

Submitted (07-MAR-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 271434)

Rat Genome Sequencing Consortium.

Direct Submission

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Submitted (13-MAY-2003) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

On May 13, 2003 this sequence version replaced gi:24819163.

The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GKNE

Center clone name: CH230-197C21

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 210037 bases at least Q40

Consensus quality: 214012 bases at least Q30

Consensus quality: 216934 bases at least Q20

Estimated insert size: 218189; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length  
 \* ([see http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 2 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 269670: contig of 269670 bp in length  
 \* 269671 269770: gap of unknown length  
 \* 269771 271434: contig of 1664 bp in length.

FEATURES  
 source Location/Qualifiers  
 1. .271434  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-197C21"  
 53225..54589  
 /notes="wgs contig"  
 73040..74665  
 /notes="wgs contig"  
 150413..151781  
 /notes="wgs contig"  
 BASE COUNT 56784 a 53038 c 51445 g 57872 t 52295 others

ORIGIN  
 56784 a 53038 c 51445 g 57872 t 52295 others

Query Match 73.1%; Score 21.2; DB 2; Length 271434;  
 Best Local Similarity 88.5%; Pred. No. 1.1e+02;  
 Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 TGAAGTTGACAGATCCAGGTGAAG 29

DB 142984 TGAAGATGGAGGATCCAGGTGAAG 142959

RESULT 14  
 BT002222/c  
 LOCUS Arabidopsis thaliana AT5g11720/T22P22\_110 mRNA, complete cds.  
 DEFINITION Arabidopsis thaliana AT5g11720/T22P22\_110 mRNA, complete cds.  
 ACCESSION BT002222  
 VERSION BT002222.1 GI:25090118  
 KEYWORDS FLI CDNA.

SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 2709)

REFERENCE  
 AUTHORS Cheuk,R., Chen,H., Kim,C.J., Shinn,P., Bowser,L., Carninci,P.,  
 Chan,M.M., Chang,C.H., Dale,J.M., Hayashizaki,Y., Hsuan,V.W.,  
 Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,  
 Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M.,  
 Palm,C.C., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A.,  
 Tang,C.C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yamada,K.,  
 Yu,G., Yuan,S., Shinozaki,K., Davis,R.W., Theologis,A. and  
 Ecker,J.R.

TITLE Arabidopsis ORF clones

JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 2709)  
 AUTHORS Cheuk,R., Chen,H., Kim,C.J., Shinn,P., Bowser,L., Carninci,P.,  
 Chan,M.M., Chang,C.H., Dale,J.M., Hayashizaki,Y., Hsuan,V.W.,  
 Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,  
 Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M.,  
 Palm,C.C., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A.,  
 Tang,C.C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yamada,K.,  
 Yu,G., Yuan,S., Shinozaki,K., Davis,R.W., Theologis,A. and  
 Ecker,J.R.

TITLE Direct Submission  
 JOURNAL Submitted (19-NOV-2002) Salk Institute Genomic Analysis Laboratory  
 (SiGAL), Plant Biology Laboratory, The Salk Institute for  
 Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,  
 USA

COMMENT  
 RIKEN Genomic Sciences Center (GSC) members carried out the  
 collection and clustering of RAFL cDNAs (RAFL CDNA : RIKEN  
 Arabidopsis Full-length cDNA ) : Seki,M., Narusaka,M., Ishida,J.,  
 Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,  
 Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PGECC (SSP) Consortium members constructed and  
 sequenced the PUN1 (ORF) clones using the RAFL cDNAs: Cheuk,R.,  
 Chen,H., Kim,C.J., Shinn,P., Banh,J., Bowser,L., Chan,M.M.,

Chang,E., Dale,J.M., Goldsmith,A.D., Jones,T., Karlin-Neumann,G.,  
 Lam,B., Lee,J.M., Lin,J., Miranda,M., Nguyen,M., Onodera,C.S.,  
 Palm,C.C., Quach,H.L., Southwick,A., Tang,C.C., Toriumi,M.,  
 Wu,H.C., Yamada,K., Yamamura,Y., Yu,G., Davis,R.W., Theologis,A.,  
 and Ecker,J.R.

Cheuk,R. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to  
 this work. Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk)  
 contributed equally to this work as PIs.

FEATURES  
 source Location/Qualifiers

1. .2709

/organism="Arabidopsis thaliana"

/mol\_type="mRNA"

/db\_xref="taxon:3702"

/chromosome="5"

/clone="UI3819"

/notes="This clone is in pUNI 51"

ecotype: Columbia

1. .2709

/notes="alpha-glucosidase 1"

/codon\_start=1

/product="At5g11720/T22P22\_110"

/protein\_id="AA072233.1"

/db\_xref="GI:25090119"

/translation="MSSLHFPNIFIVVVVFLSRSSQVVEEESTVVGYYVVRVSV

GVDSNRQVLTAKDLIKPSSVYAPDIKSLNLHVSLETSELRIRITDSSQORSEIDP

VIPRAGNHSPPRESTDGSGSPENNFLADPSSDLVFTLHNTTPGFSVSRSSGDL

FDTSPPSSDSNTYIFPKDQLQSSALPENRNLXIGIEHTKXSFRLIPGETWLNWA

DIGSEPNLVGSHFFPYMDVRGKNEAGTTTHGVLLNSGMDVYEGREITVYV

GGVIDVYFAGSPPEMNQYTELIGRAPMYPNFGHCQYKQVSKVSLVYVDPG

AKAGIPLEVMTDIDYMGYKDFLPVNFPEDMQSFVDTLHNGKQKYLILDPGIG

VDSSTYTNRGMEADYIFKRNQEPYLGCEVMPGVKVPDFLNPAAATFWSNEIKMFGI

LPDGLMIDMNELSNPTSLSSGSLDDPPYKINNSGDKRPINNKTVPATSTHFGNI

SEYDAHNLGLLEAKATHQAVDITGKRPEILSRSTFVSSGKYTAHNTGNAKAWED

AYSIPGLNFGFCIPMWGADICGSHDITTEELCRWIQLGAYFPARDHSSLGTRAQ

ELYLWDSVASSARKVGLRMLRPLHLYTLMEAHVSGNPARTPLFFSPDPTKTEID

SOFLIGKISVSPALQGVAVDAYFPAGNFDLFNYSFAVGGDSKGVRLDPADHV

NVHREGSIVAMOGKALTRDARKTPQLLVASRLNISGELFLDGGENLRMGAGGG

NRWTLVKFRCYTGKSVLRSEVNPVYASKMWSIKGTVFVGFEFNVENKTVYVRT

SERLSRPSRLIKTVSDNDPRLSVEVSKLSLLVGKKFEMRLRLT"

BASE COUNT 729 a 584 c 663 g 733 t

ORIGIN

Query Match 72.4%; Score 21; DB 8; Length 2709;

Best Local Similarity 82.8%; Pred. No. 95;

Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGGTGAAGTTGACAGATCCAGGTGAAG 29

DB 1096 CCGGGAATTTACAGGATCCAAAGTGAAG 1068

RESULT 15

AT053414/c

LOCUS Arabidopsis thaliana AT5g11720/T22P22\_110 mRNA, complete cds.

DEFINITION Arabidopsis thaliana AT5g11720/T22P22\_110 mRNA, complete cds.

ACCESSION AY053414

VERSION AY053414.1 GI:15450744

KEYWORDS FLI CDNA.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 2769)

REFERENCE

AUTHORS Shinn,P., Chen,H., Cheuk,R., Kim,C.J., Koesema,E., Meyers,M.C.,  
 Banh,J., Bowser,L., Carninci,P., Dale,J.M., Goldsmith,A.D.,  
 Hayashizaki,Y., Ishida,J., Jiang,P.X., Jones,T., Kamiya,A.,  
 Karlin-Neumann,G., Kawai,J., Lam,B., Lee,J.M., Lin,J., Liu,S.X.,  
 Miranda,M., Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J.,  
 Pham,P.K., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A.,  
 Tang,C.C., Toriumi,M., Yamada,K., Yamamura,Y., Yu,G., Yu,S.,  
 Shinozaki,K., Davis,R.W., Theologis,A. and Ecker,J.R.

TITLE Arabidopsis cDNA clones  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2769)  
AUTHORS Shinn, P., Chen, H., Cheuk, R., Kim, C.-J., Koesema, E., Meyers, M.C.,  
Barth, J., Bowser, L., Carninci, P., Dale, J.M., Goldsmith, A.D.,  
Hayashizaki, Y., Ishida, J., Jiang, P.X., Jones, T., Kamiya, A.,  
Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J.M., Lin, J., Liu, S.X.,  
Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J.,  
Pham, P.K., Quach, H.B., Sakurai, T., Satou, M., Seki, M., Southwick, A.,  
Tang, C.C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S.,  
Shinozaki, K., Davis, R.W., Theologis, A. and Ecker, J.R.

TITLE Direct Submission  
JOURNAL Submitted (28-AUG-2001) Salk Institute Genomic Analysis Laboratory  
(SIGNAL), Plant Biology Laboratory, The Salk Institute for  
Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,  
USA  
RIKEN Genomic Sciences Center (GSC) members carried out the  
collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN  
Arabidopsis Full-Length cDNA') : Seki, M., Narusaka, M., Ishida, J.,  
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,  
Hayashizaki, Y. and Shinozaki, K.

COMMENT The Salk, Stanford, PGSC (SSP) Consortium members carried out the  
sequencing and annotation of the RAFL cDNAs: Shinn, P., Chen, H.,  
Cheuk, R., Kim, C.-J., Koesema, E., Meyers, M.C., Barth, J., Bowser, L.,  
Dale, J.M., Goldsmith, A.D., Jiang, P.X., Jones, T., Karlin-Neumann, G.,  
Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M., Nguyen, M.,  
Onodera, C.S., Palm, C.J., Pham, P.K., Quach, H.B., Southwick, A.,  
Tang, C.C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S.,  
Davis, R.W., Theologis, A., and Ecker, J.R.

Shinn, P. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to  
this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk)  
contributed equally to this work as PIs.

FEATURES  
source

1..2769  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/db\_xref="taxon:3702"  
/chromosome="5"  
/clone="RAFL09-14-L23(R13819)"  
/note="ecotype: Columbia"  
1..26  
27..2735  
/note="alpha-glucosidase 1"  
/codon\_start=1  
/product="AT5g11720/T22P22.110"  
/protein\_id="AAK96644.1"  
/db\_xref="GI:15450745"

5'UTR  
CDS

/translation="MSSLHWFPIFVIVVFFSLRSQVLEEEESTVGYGVVRSV  
GVDSNRQVLTAKDLIPSSVYADPIKSLNLHVSLETSELRIRITDSSQORWEIPE  
VIPRAGNHSRPFSTEEDGNSPENFLADPSDLVFTLHNTTTFGFSVSRSSGDIL  
FDTSPDSDSTYFIKQDFQLQSSALPENRSLNYGIGENTKRSFRLIPGETMTLWNA  
DIGENPDVNLVYGGHPVMDVGRSGNEAGTTHGVLNLSNGMDVKYEGHRTYNNV  
GKIDLVYFAGPSPEMVMNOYTELIGRPADPYWSFGFHOCYKYNVDELYVVDGY  
AKAGIPLEVMTDIDYMGYKDFLDPNPEPDKMQSFVDTLHNGQKXYVLLDPGIG  
VDSSTGTNRGEADYFIKRNGEFYLGVEVMPGKVIFFDFLNPAAITPWSNEIKMFQEI  
LPDLGLWDMELNFIITPLSSGSLDDPPYKINNSGDKRPINNTKVPATSIHFQNI  
SEYDAHNLVGLLEAKATHQAVDITGKRFILSRSTFVSSGKYTAHWGDAKAWEDL  
AYSTPGLNFGILFGIPMYGADICGFSHTTEELCRWIOLCAFYFPFARDHSSLGTAQ  
ELYLWDSVASSARKVGLRMLLPHLYTLMYEAHVGNPIARPLFFSFPODTKTYEID  
SQFLGKISVPALVKCAVADAYFAGNWFELFNYSEAVGDSGKHVLTDPADHV  
NVHVEGSIIVAQGEALTTRARKTPQLLVASRLNISGELFDGGENLRMGAGG  
NRDWTLVKFRCTVLRSEVNVNPEVNSKMKWSIGKTVTFVGFEVNVNKTVEVRT  
SERLSRPSRLIKTVTSDNDPRFLSVESKLSLLVGGKFMRLRLT"  
2736..2769

BASE COUNT 763 a 590 c 671 g 745 t  
ORIGIN

Query Match 72.4% Score 21; DB 8; Length 2769;  
Best Local Similarity 82.8% Pred. No. 95;  
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGGTGAAGTTGACAGGATCCCAAGGTGAAG 29  
DB 1122 CCGGGAATTTACAGGATCCCAAGGTGAAG 1094

Search completed: October 27, 2003, 11:10:33  
Job time : 2688 secs